

• EXSTRESS PROPERTY OF THE TRANSPORT OF THE PROPERTY OF THE PR

OTTIC FILE COPY

COMPUTER SYSTEMS LABORATORY





STANFORD UNIVERSITY · STANFORD, CA 94305-2192

Research in VLSI Systems

Technical Progress Report

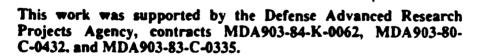
April 1986 - December 1986

Computer Systems Laboratory Integrated Circuits Laboratory Center for Integrated Systems

> CLEARED FOR OPEN PUBLICATION

APR 0 8 1987

IRECTOHALE FUR FREEDOM OF INFORMATION AND SECURITY REVIEW (OASD-PA) DEPARTMENT OF DEFENSE



The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

87 1384

Research in VLSI Systems

Progress Report for April 1986 - December 1986

Center for Integrated Systems
Stanford University
Stanford, California 94305

General Purpose VLSI-Based Multiprocessors
DARPA Contract No. MDA903-83-C-0335
DARPA Order No. 3773-6
Principal Investigator: John Hennessy
Computer Systems Laboratory
Monitored by W. Bandy

A Fast Turn Around Facility for Very Large Scale Integration (VLSI)
DARPA Contract Nos. MDA903-80-C-0432 and MDA903-84-K-0062
Principal Investigator: James Plummer
Integrated Circuits Laboratory
Monitored by W. Bandy

Out Protect to

Atter enfice

Table of Contents		
1 Design Description, Analysis, and Synthesis	4	
1.1 Circuit Modeling for Simulation	4	
1.2 Final Layout Checks	4	
1.3 Functional Simulation	5	
2 VLSI Processor Architecture and Software	6	
2.1 MIPS-X: A High Performance VLSI Computer	6	
2.1.1 Hardware Status	6	
2.1.2 Cache Studies	7	
2.1.3 Making LISP run fast	8	
2.1.4 Reducing the Cost of Branches	9	
2.1.5 MIPS-X Summary	10	
2.2 Multiprocessor Support for MIPS-XMP	10	
2.2.1 Decomposing Parallel Programs	10	
2.2.2 MIPS-XMP Summary	11	
2.3 Tester Memory	11	
3 Testing	11 13 13 13 14	
3.1 Tester Memory	13	
3.2 Testable CMOS Design	13	
3.3 Parametric Testing and Diagnostics	14	
3.3.1 In-process Monitors	14	
3.3.2 End-of-process Monitors	15	
3.3.3 Supporting Activities	18	
3.3.4 Wafer Processing and Testing	15 18 19	
3.3.5 Electrical Alignment Test Structures	20	
4 Fast Turn-Around Laboratory	27	
4.1 Fable: Knowledge Representation for Semiconductor	27	
Manufacturing		
4.1.1 Lessons	27	
4.1.2 Knowledge and Representation	38	
4.1.3 Knowledge Tools	38	
4.1.4 Applications Projects	38	
4.1.5 Object Oriented SECS-I/II Equipment Communications	39	
4.1.6 Collaboration	39	
4.1.7 New Automation Course	40	
4.1.8 New Electronic Discussion Group	40	
4.2 Microlithography	43	
4.2.1 Electron Beam Lithography	43	
4.2.2 Defect Inspection	44	
4.2.3 Langmuir-Blodgett Films	44	
4.3 Processes, Devices, and Circuits	45	
4.3.1 Interfacial Charges and Time-Dependent Breakdown in		
MOS Devices with Rapidly Grown Ultrathin SiO ₂ Gate		
Insulators		
4.3.2 Dry Etching	47	
4.4 Interconnections and Contacts	50	
4.4.1 Objective	50	

Abstract

This report summarizes progress in the DARPA funded VLSI Systems Research Projects, from April 1986 to December 1986, inclusive: The major areas under investigation have included: analysis and synthesis design aids, applications of VLSI, special purpose chip design, VLSI computer architectures, reliability studies, hardware specification and verification, and VLSI fabrication. The major research problems are introduced and progress is discussed; the Appendix contains a list of published research papers from these projects.

Key Words and Phrases: VLSI, design automation, computer-aided design, special purpose chips; VLSI computer architecture, routing, layout, memory reliability, IG fabrication.

Same not below the

This work was supported by the Defense Advanced Research Projects Agency, contracts MDA903-84-K-0062, MDA903-80-C-0432, and MDA903-83-C-0335.

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

Executive Summary

The major progress of note for this period is as follows:

- 1. MIPS-X: a very high performance VLSI processor. MIPS-X [Chow 86, Chow 85, Horowitz 87] a project to develop a very high performance processor to be used as the node processor in a high performance multiprocessor. Like MIPS,:MIPS-X uses a simplified instruction set, a deep pipeline, and code reorganization to increase performance. Unlike MIPS, MIPS-X contains an on chip instruction cache, and supports both coprocessor and a multiprocessor environment. The chip was submitted to fabrication in May and we received working chips in October. Preliminary performance results indicate a clock speed of 17 MHz (with a target of 20 MHz). System testing and integration remain to be done. The chip was designed in a 2µ two level metal technology, and we expect to shrink to 1.25 µ. Several supporting research projects showed major progress, including: work in performance analysis and performance estimation for large caches, and studies of branch prediction techniques, development of LISP compilers, and new compiler optimization algorithms.
- 2. Cache Studies. To support MIPS-X and MIPS-X-MP, very large caches are required. Such caches can maintain information across operating system calls and process switches. Classical trace data does not adequately drive very large caches, nor usually contain multiple processes. A technique for obtaining such data has been developed¹. The resulting traces are leading to new insights on the usefulness of very large caches [Agarwal 86a, Agarwal 86b], and the traces have been made available to other groups in the university research community. Using this data, a highly accurate model that accommodates many cache design parameters has been developed.
- 3. Software support for RISC processors. We have continued to explore methods of improving the effective performance of a processor by improving the quality of the code generated by the software system. This effort involves work both in optimizing compilers [Chow 83, ChowHenn 84], and code scheduling [Gross 83, McFarling 86]. We are also looking at the performance of the MIPS-X architecture for the LISP language [Steenkiste 86], by creating a version of PSL for the machine.
- 4. Automatic partitioning of parallel programs. A system for partitioning dataflow graphs into multiple tasks for execution on a parallel processor has been developed. The first version concentrates on a model using compile-time partitioning and scheduling [Sarkar 86a]. We have also developed a model for dynamic scheduling [Sarkar 86b]. Related work has focused on optimization problems in the functional languages that generate our data flow graphs.
- 5. RSIM. We have extensively changed the models that the switch level simulator, RSIM, uses to determine the value of a node. These changes involve changing the basic transistor models to better approximate the nonlinear transistor characteristic, the timing models to include the effect of input slope and distributed RC networks, and the charge sharing model to include the effect of resistance on charge

Jointly with Dig. al Equipment Corporation.

sharing [Chu 86].

- 6. Testing Chip In conjunction with MOSIS, we have started the design of a special purpose memory chip that will enable us to build a high speed tester at a low cost [Miyamoto 87]. The chip acts as a small test vector memory and a set of very flexible input output pads. Each chip drives 16 DUT pins, and is housed in a 84 PGA package. The first versions of this chip have been fabricated and are fully functional.
- 7. Testable CMOS Design. Design-for-testability (DFT) techniques have been developed to improve the testability of static CMOS circuits. These techniques are used to design fully-testable combinational circuits. Conventional gate-level automatic test pattern generators (ATPGs) instead of the less efficient switch-level ATPGs can be used to generate tests for switch-level faults in this type of circuit.
- 8. Computer Support Fable. We have initiated a course entitled "Automation of Semiconductor Manufacturing" which is bringing together Al and wafer fabrication experts to attack several problems of importance to the Computer Automated Fabrication effort. These groups are working in an advanced TI Explorer/KEE environment.
- 9. Computer Integrated Manufacturing e-mail discussion group. A moderated interuniversity news group has been established to discuss matters of interest to the Computer Automated Fabrication community. Join by sending your net address to IC-CIM-Request@Sierra.Stanford.EDU
- 10. Electrical alignment test structures. A comprehensive set of test structures which monitor δX and δY registration accuracy have been developed.
- 11. Template-set matching for random defect detection. A 2 µm CMOS circuit has been designed to aid in random defect inspection of masks and integrated circuits. A template-set matching scheme has been applied to the task of defect detection and, more recently, to defect classification.

Technical Progress

1 Design Description, Analysis, and Synthesis

1.1 Circuit Modeling for Simulation

We have continued our work on improving the models that are used in switch level simulation. Our work in this area is based on the RSIM simulator from MIT. We have used the basic event scheduling engine and the same user interface, but have made extensive changes to the new value evaluation models. Our original changes were necessitated by the MIPS-X design. RSIM in its original form would not correctly model the circuits used in that design, nor were its timing models accurate enough. Recently, we have concentrated on providing a better model for charge sharing in MOS circuits. Although charge sharing has a very important effect in circuits, most simulators use an ad-hoc approach to its modeling. We have found a method of using a 2 time constant model of a circuit to provide a natural method of modeling charge sharing. The first time constant represents the charge sharing event and the second time constant represents the driven response. We have used this method to model circuits that RSIM previously failed to handle. We are now looking at completely rewriting the circuit evaluator in RSIM to correct its remaining problems.

Staff: C.Y. Chu, M Horowitz

References: [Chu 86]

1.2 Final Layout Checks

Before MIPS-X was 'taped-out' we felt a need to check the final version of the layout for certain errors that the Magic layout system does not check. These checks include looking for floating well, zener diodes (well/substrate plug butting into diffusion not connected to the supply) and resistance extraction. These checks are a result of collecting horror stories from other designers and then figuring out a way to make sure that you don't fall into the same trap. For example, the problem with zeners was one that we had not considered until a designer of another chip and members of the MIPS-X team traded war stories. When a check for zeners was constructed, we found a number of sections of MIPS-X that might have failed because of this problem.

We were able to use the Magic system to check for both floating wells and zener diodes by using the extractor in a way that was probably never intended. The well check is very simple. The key is that the node numbers that Magic generates contains the plane where the object is resides. Since well is on a separate plane, all one needs to do is make well have a very large area capacitance (to insure that it is put into the sim file) and then look in the flat sim file for nodes that are on the well layer. The path name of this node will give the location of the floating node. If the node is connected to a power supply the node name will be aliased to Vdd or Gnd. Zener diodes can be found in a similar manner. The circuit needs to be exacted twice, once with plug connected to diff, and once with it not connected. Diffing the '.ext' files will point out potential zeners without flagging correct abutment of well plug with a Vdd or Gnd contact.

Resistance extraction posed a more difficult problem, since Magic really does not have true

resistance extraction. We have integrated a resistance extractor into Magic and have run it on the MIPS-X database. The resistance extractor uses the original Magic resistance values as a filter to remove nodes that cannot have a significant resistance value. The resistance of the remaining nodes is exacted, and nodes with a significant resistance are partitioned and the new network is added to the sim file. We are currently working on extracting power and ground nets and using this information to estimate power and gnd noise in a VLSI circuit.

In the final stages of debugging MIPS-X, we found a need to run simulator from Magic, and allow the user to point at layout and find the value of that node in the simulation. During this past period an interface to the RSIM simulator was added to Magic, making this operation possible. Although this interface came up too late to be used in MIPS-X it has been extremely helpful in other chips that were under design during that time. This included a Tester Memory and high speed Mult and Div chips. We are now working on an extension to this interface that will allow a user to modify the layout to correct a bug, incrementally extract this change, and load the change into the simulator.

Staff: D. Stark, M Horowitz, M. Chow

Related Effort: Magic at Berkeley

1.3 Functional Simulation

In the MIPS-X design process, a disposable functional simulator was written because a general purpose functional simulator was unavailable. Correcting this tool deficiency, we've taken CSIM (Univ of Colorado) and made substantial modifications to improve error reporting, human interface, design partition techniques, and modeling capability. We are supporting this simulator for the Stanford community for new designs and classes. In conjunction with U of Col., we jointly plan to release a new version incorporating our changes in January of 1987.

Besides the CSIM support activities, the functional simulator is being used as the starting platform for research in incremental simulation, tool integration, and parallel simulation studies. The incremental simulator development has progressed so that both components and wires can be added or deleted. This work has only entered the first testing phase. This prototype has not investigated compression techniques on the internal state store which is required for a completed program. An incremental net-list flattener (also generates change tokens) is nearing completion. Both simulator and flattener need to be completed before significant testing can occur.

The simulation integration effort involves the programmatic connection between CSIM and other simulators (e.g. RSIM) or physical test equipment (e.g. medium tester or logic analyzer). The motivation is that the test vectors and display environment can be common among the various tools and more importantly, that signal interfaces can be automatically verified. The first phase will connect CSIM and RSIM. Second, will be the addition of a physical tester. This work has only recently begun, but a small prototype should be completed for use in January of 1987.

The parallel simulation research first investigated both performance implications and potential parallelism when modeling the same circuit at different abstraction levels (instruction, behavioral, RTL, and gate). Roughly, the performance decreases by a decade and the component count increases by a decade going from the high to the low level. The observed parallelism was

approximately .1% per clock tick and 20% per clock cycle. The limited available parallelism suggests the future research directions into simulation pipelining, unit delay simulation, and chaotic time.

Staff: B. Alverson, S. Y. Hwang, L. Soule, T. Rokicki, K.Y. Choi, T. Blank

Related Effort: CSIM Univ of Colorado

2 VLSI Processor Architecture and Software

2.1 MIPS-X: A High Performance VLSI Computer

The MIPS-X uniprocessor design goal is a machine with a 20 MIP peak instruction rate, and an 'average' throughput of over 10 MIPs. The architecture is of the reduced instruction set variety, but also ventures into two new and important areas:

- 1. supporting high performance co-processors, and
- 2. providing the capability to be used in a medium-scale multiprocessor environment.

In addition, we have several closely related activities. These involve studying the implementation of LISP on MIPS-X, and the performance and analysis of very large caches.

2.1.1 Hardware Status

During this period, we completed the design of the MIPS-X processor, and submitted the chip for fabrication. The completed design was debugged using the functional simulator to generate vectors for a switch level simulator (RSIM) running on an extracted circuit description. The simulator ran at a rate of about 1 clock cycle per uVax CPU minute, or 1400 cycles/uVax-day. At the end of the design, we typically kept 4 uVax CPU bound 24 hours a day. We succesfully ran several short programs, including a set of diagnostics written to test out tricky instruction sequences. We also ran the programs setting exception high at random times to test the interrupt and exception hardware. The net result was the machine successfully ran about 30 - 40K cycles of test programs before we submitted it for fabrication.

During this extensive testing we also ran a timing verifier over the design to look for slow paths in the machine. Although we found some paths that would keep the machine from hitting its target speed of 20MHz, there were not paths that would prevent the machine from running at 10 - 12MHz. We felt that this performance on first silicon was acceptable, and it was not worth risking the functionality of the part to improve performance. In addition, since all the timing simulation had been done with worst-case numbers, there was some hope that the real chips would run faster than the simulations.

The chips were sent to MOSIS for fabrication on the first 2µ CMOS run that closed around May 1. This run was then delayed and sat at MOSIS for about 3 months before being sent to VTI. In the meantime Xerox Corporation offered to fab MIPS-X in one of their VTI 2µ CMOS runs. This run was also delayed, but made it out of fab on the beginning of October. The MOSIS run also came out of fabrication in October, but because of a mask manufacture error the silicon was non functional. During this past month we have extensively tested the processor. Within 48 hours of receiving the part, we had test data that indicated that the processor was functional.

After working on fixing the testing software and integrating the functional simulator to the tester, we were able to test the processor at low speeds. The processor was completely functional. The only design error discovered was a shorted line in the instruction cache, an error not caught because the cache was not switch level simulated. Actually, the error was not caught because we forgot to examine the output of the internal cache drivers. We actually simulated the faulty circuit; we simply did not look at the right outputs.

We have also run some preliminary speed tests. Using Pico-probes, we have measured waveforms of some internal signals. These results indicate the top machine speed will be around 60ns cycles, or about 17MHz. We were then able to load a simple program into the cache and have the machine run the program at 17 MHz. We cannot do more extensive speed testing since we do not have any equipment capable of generating test vectors at a fast enough rate. To further exercise the chip, we are building a test board that will plug into a VME bus. This board will allow us to run exercise the processor at higher speeds. The board is being simulated with CSIM (see the CAD tool section) and will be set for fabrication by the first of the year.

We have corrected the error on the processor and resubmitted the chip for fabrication. We have also begun changing the processor to fix the slow paths in the machine, and hope to send a new revision of the the chip early in the first quarter of next year. We are also ramping up the design effort on the cache controller chip for MIPS-X, and expect to submit this chip for fabrication during the second quarter of 87.

2.1.2 Cache Studies

The increasing performance demanded of caches in current high-speed computer systems requires that our analysis and prediction of cache performance become more exact. Unfortunately, current cache research has not been able to do so, largely because of the unavailability of efficient analysis techniques for large caches and the difficulty of data collection for realistic operating system and multitasking environments. This research addresses these problems.

We developed a new tracing method called ATUM [Agarwal 86a] that turned out to be extremely helpful in generating realistic numbers for cache hit rates. This work was done jointly with, and partially supported by, Digital Equipment Corporation. The method uses the microcode of a running system to record the address of every memory reference that the machine generates. This trace is complete in the sense that it includes both user and kernel references, and contains information about context switches and interrupt activity. The traces have been distributed to various groups in the academic research community.

We investigated several efficient cache analysis techniques, including: a mathematical cache model [Agarwal 86b], a trace sampling scheme, and a trace compaction method called cache filtering with blocking. The cache model uses a few parameters extracted from the address trace as inputs and gives miss rate as a function of cache size, set size, block size, and multiprogramming level. Validations against the ATUM traces showed the predicted values to be similar to the results of trace driven simulation while requiring very little calculation time. The trace sampling scheme together with an understanding of transient cache behavior allows accurate empirical estimation of steady-state cache miss rates from short trace samples, thereby significantly reducing simulation time. If sampling is used in conjunction with our trace compaction technique, the potential for increasing simulation efficiency is enormous, albeit at

some loss in accuracy.

Accurately characterizing cache behavior using our sampling methodology and the distortion-free ATUM traces shows that both operating system and multiprogramming activity significantly degrade cache performance, with an even greater proportional impact on large caches. We have found that although system references are only 10 - 50 percent of the total references (with an average of about 20%) their intrinsic miss rate is so high it causes the miss rate for the user/system combination to double from a user only reference stream. Similarly, the large combined working sets of multiple processes increases the miss rate for multiprogrammed caches. Sharing system references across all processes decreases this penalty.

Our studies show that large cache performance is highly sensitive to the technique adopted to manage the cache in a multiprogrammed environment. Virtual-address caches with process identifiers and physically-addressed caches have the highest hit rates. The dismal performance of cache flushing on a context switch — currently a popular technique for small, virtual cache management — limits its usefullness for large virtual caches.

We took a second look at associativity. For large caches an associativity of two picked up almost all the benefits of full associativity, and doubling the associativity seldom had a significantly better hit rate than doubling the cache size. Furthermore, using average access time as a metric of cache effectiveness, the advantage of increasing associativity trades off against an increase in cache access time. Therefore, a large direct-mapped cache with some simple enhancements, (like hashing), is likely to outperform complex set-associative organizations. This only reaffirms our faith in simple implementations for best overall performance.

2.1.3 Making LISP run fast

The high-level language LISP has some features, like runtime type checking, that make it very different from C and Pascal, the two languages focused on in the design of MIPS-X. To determine which LISP operations are time critical, data is needed that characterize the execution behavior of LISP programs. This data will also allow us to evaluate the MIPS-X architecture as a host for the execution of LISP programs.

We ported the Portable Standard Lisp compiler to the MIPS-X processor, and collected dynamic profiling information for 11 LISP programs [Steenkiste 86]. These measurements showed that the instruction frequencies for LISP are similar to those for Pascal. Two important differences are that LISP programs execute significantly more procedure calls, and that the group of alu instructions in LISP is dominated by bitfield operations used for tag handling, but by arithmetic instructions in Pascal. We found that almost three fourths of the program execution time is used for three operations: tag handling (23%), procedure calls (26%) and stack accesses (22%). We looked at optimizations for each of these 3 time consuming operations.

Each data object in LISP has a tag that contains its type. On general purpose architectures, type checking and using the data often requires a bitfield operation to separate the tag and the data part. By choosing the tags so that some of these operations are eliminated, our LISP programs ran 5% faster. The high cost of runtime type checking has encouraged others to use special hardware to support tags (for example in LISP machines), and our measurements show that a moderate amount of hardware support could speedup our LISP programs between 10% and 20%. The exact speedup will depend on how much runtime type checking can be eliminated by a

compiler that tries to derive the type of objects at runtime, possibly using declarations provided by the user.

To reduced the cost of procedure calls, we first optimized and inlined a number of time critical primitive LISP operations. This speeded up the programs about 16% — half of this gain is the results of eliminated procedure calls. In a next step, we merged small procedures under control of the compiler. A study of the effect of merging on the miss rate in the MIPS-X on-chip instruction cache showed that aggressive merging actually slows down programs because the increase in code size also increases the instruction miss rate. By only merging small, non-recursive procedures, this effect could be reduced, and we measured an overall speedup of 6% as the result of merging user functions.

The high procedure call frequency in LISP makes effective per-procedure register allocation less effective than in a C or Pascal environment. We implemented a simple inter-procedural register allocator that propagates information about register usage in the program call graph (similar in spirit to Wall's approach). As a result, different procedures use different registers, so fewer registers have to be saved across procedure calls. This allowed us to eliminate 70% of the stack accesses, and the 11 LISP programs ran an average of 10% faster. Recursion was the limiting factor on the performance of the inter-procedural register allocator. Measurements of the effect of register windows on the memory traffic showed that a register file with at least 80 registers would be required to eliminate the same number of stack accesses as inter-procedural register allocation.

The performance results of MIPS-X for LISP look very encouraging. Although MIPS-X does not have any tagging hardware, it does have sufficient support for bitfields to handle tags efficiently. The execution of the Gabriel benchmarks on the MIPS-X simulator, which includes the effect of the (off chip) cache, show a performance that is significantly higher than the Symbolics 3600 LISP machine. Full runtime type checking was used for these simulations, and none of the above optimizations were included.

2.1.4 Reducing the Cost of Branches

Pipelining is the major organizational technique that computers use to reach higher single-processor performance. A fundamental disadvantage of pipelining is the loss incurred due to branches that require stalling or flushing the pipeline. If nothing special is done, branches interfere with the normal pipeline because the following instruction depends on a condition evaluation and perhaps the fetch of an non-sequential target.

Techniques for speeding up branches can be divided into those that stress hardware and software. Special hardware can predict from past behavior which direction a branch will go in the future. Also, the target instruction can be stored in a Branch Target Buffer where if the prediction is correct, the target instruction is immediately available. In software, branches can be sped up by trying to schedule instructions after a "delayed branch" which always executes the one or two instructions after a branch.

In the design of the MIPS-X processor, two new techniques were developed that provide very fast branches with minimal hardware overhead. First, MIPS-X squashing branches execute the two next instructions only if the branch is taken. This allows the slots to always be filled, unlike delayed branch slots which can only be filled with instructions that can always be executed

whether the branch is taken or not. The second technique is to use profile information from a previous run of the program to drive branch scheduling. Profile information is as accurate as hardware prediction but does not require any hardware. Given a profile, the compiler can handle branches differently if they are usually not taken.

To evaluate these new techniques, they were simulated on a set of pascal benchmarks for a machine with an inherent 2-cycle branch delay, like MIPS-X. As the table below shows, a Profiled Squashing Branch performs better than a hardware intensive Branch Target Buffer without the need for extensive hardware.

	Cycles/ Branch	/ Machine Performance
Simple Branch	3.00	1.25
Branch Target Buffer	1.32	1.04
Delayed Branch	2.21	1.15
Profiled Squashing Branch	1.27	1.03
"Ideal" Branch	1.00	1.00

2.1.5 MIPS-X Summary

Staff: S. Przybylski, C. Y. Chu, J. Hennessy, M. Horowitz, M. Wing, P. Chow, J. Acken, A. Agarwal, S. Richardson, S. McFarling, M. Ganapathi, D. Stark, R. Simoni, S. Tjiang, P. Steenkiste

Related Efforts: SPUR (Berkeley)

References: [Hennessy 84], [Chow 86], [Agarwal 86a, Agarwal 86b], [Agarwal 87], [Steenkiste 86], [Horowitz 87]

2.2 Multiprocessor Support for MIPS-XMP

Our work on caches supports the MIPS-X design, but it is even more critical for our multiprocessor activities. To date the architectural work for MIPS-XMP has focused on high performance memory hierarchies needed to support 8-10 15-mips processors. We are also making progress on our software activities as described below.

2.2.1 Decomposing Parallel Programs

There are three fundamental problems to be solved in the execution of a parallel program on a multiprocessor - identifying the parallelism in the program, partitioning the program into tasks and scheduling the tasks on processors. Whereas the problem of identifying parallelism is a programming language issue, the partitioning and scheduling problems are intimately related to the number of processors and the synchronization and communication overhead in the target multiprocessor. It is desirable for the partitioning and scheduling to be performed automatically, so that the same parallel program can execute efficiently on different multiprocessors. We have investigated two solutions to the partitioning and scheduling problems. The first approach is based on a macro-dataflow model [Sarkar 86b], where the program is partitioned into tasks at compile-time and the tasks are scheduled on processors at run-time. The second approach is

CHARLES AND A COMPANY OF THE PROPERTY OF THE P

based on a compile-time scheduling model [Sarkar 86a], where the partitioning of the program and the scheduling of tasks on processors are both performed at compile-time.

Both approaches have been implemented to partition programs written in the single-assignment language, SISAL. The inputs to the partitioning and scheduling algorithms are a graphical representation of the program and a list of parameters describing the target multiprocessor. Execution profile information is used to derive compile-time estimates of execution times and data sizes in the program. Both the macro-dataflow and compile-time scheduling problems are expressed as optimization problems, which are proved to be NP-complete in the strong sense. We present approximation algorithms for these problems. The effectiveness of the partitioning and scheduling algorithms is studied by multiprocessor simulations of various benchmark programs for different target multiprocessor parameters.

As mentioned above, both the partitioner for macro-dataflow and the partitioner-cum-scheduler for compile-time scheduling have already been implemented to partition SISAL programs. The partitioning is actually performed at the level of SISAL's graphical intermediate form, IF1. We extended the Livermore IF1 interpreter to produce trace files for multiprocessor simulations. We have a variety of SISAL benchmark programs, from small programs like Matrix Multiplication, Merge-exchange Sort, FFT (approximately 100 lines each) to larger programs like SIMPLE and SLAB (approximately 2000 lines each).

The goal of our project is to make single-assignment languages like SAL and SISAL run efficiently on real multiprocessors. The following additional pieces need to come together to build a complete compiler system:

- 1. A code generator for single-assignment languages. This is a hard problem to solve completely; we are pursuing a general solution and also a straightforward SISAL-to-C translation, which can already generate code for small benchmark programs.
- 2. Synchronization primitives for compile-time scheduling.
- 3. Run-time scheduler for macro-dataflow.
- 4. Experiments on Encore, NCUBE, and a workstation cluster.

This work is also partially support by an NSF PYI award.

2.2.2 MIPS-XMP Summary

Staff: S. Przybylski, J. Hennessy, M. Horowitz, M. Wing, P. Chow, A. Agarwal, J. Celoni, V. Sarkar, K. Gopinath, H. Davis, K. Gharachloo, S. Tjiang, J. Rose

Related Efforts: SPUR (Berkeley), Butterfly (BBN), Cosmic Cube (Caltech), RP3 (IBM)

References: [Sarkar 86a], [Sarkar 86b], [Hennessy 86]

2.3 Tester Memory

The tester memory is an attempt to use VLSI technology to make VLSI chips easier to test. The Data Generator-Receiver chip really serves two functions: it acts as a small high speed vector memory, allowing burst vector rate of over 20MHz, and it acts as a configurable set of input/output pads optimized for driving the DUT (device under test). The current version of the

DGR stores 256 vectors per pin, contains the electronics for 16 DUT pins, and is housed in a 84 pin PGA.

During this period, we have completed the design and verification of 2 versions of the DGR chip. The first version was designed in a 3μ CMOS technology. This chip was submitted to MOSIS for fabication in June and we received working silicon in September. This silicon was extensively tested, first using the Sun Kit 1, a version of the Stanford Medium Tester distributed by Bob Parker's group at ISI, and then by using a specially designed test fixture that we built for speed testing. The chips are completely functional, and run at over 10MHz. During the testing of these chips we refined some of the features of the chip. These refinements were added to the 2μ version of the chip that was submitted in October. The 2μ version of the chip should be able to sustain a vector rate of roughly 20MHz.

We are now working closely with Bob Parker to integrate these parts into a new low cost tester to replace the Sun Kit 1 tester. Two types of testers are invisioned. One tester would stress low cost. It would be directed at providing a low cost method of testing the chips fabricated on MOSIS class runs. The other tester is meant as a replacement of the current generation tester. It will provide a more complete tester interface, for example, providing true bidirectional DUT pads, and also providing a limited high speed test capability. The initial specification of the new tester has begun, and we hope to have a working prototype in about a year.

We have also begun work on a next generation set of tester chips. The goal of this project is to develop pin drive electronics that can set edges to about a 2ns resolution, and operate at a vector rate of 30-40MHz. We have finished the preliminary design of the pin drive electronics and hope to have a test chip out by the middle of next year.

Staff: M. Horowitz, J. Miyamoto, J. Gasbarro

References: [Miyamoto 87]

3 Testing

3.1 Tester Memory

The tester memory is an attempt to use VLSI technology to make VLSI chips easier to test. The Data Generator-Receiver chip really serves two functions: it acts as a small high speed vector memory, allowing burst vector rate of over 20MHz, and it acts as a configurable set of input/output pads optimized for driving the DUT (device under test). The current version of the DGR stores 256 vectors per pin, contains the electronics for 16 DUT pins, and is housed in a 84 pin PGA.

During this period, we have completed the design and verification of 2 versions of the DGR chip. The first version was designed in a 3μ CMOS technology. This chip was submitted to MOSIS for fabication in June and we received working silicon in September. This silicon was extensively tested, first using the Sun Kit 1, a version of the Stanford Medium Tester distributed by Bob Parker's group at ISI, and then by using a specially designed test fixture that we built for speed testing. The chips are completely functional, and run at over 10MHz. During the testing of these chips we refined some of the features of the chip. These refinements were added to the 2μ version of the chip that was submitted in October. The 2μ version of the chip should be able to sustain a vector rate of roughly 20MHz.

We are now working closely with Bob Parker to integrate these parts into a new low cost tester to replace the Sun Kit 1 tester. Two types of testers are invisioned. One tester would stress low cost. It would be directed at providing a low cost method of testing the chips fabricated on MOSIS class runs. The other tester is meant as a replacement of the current generation tester. It will provide a more complete tester interface, for example, providing true bidirectional DUT pads, and also providing a limited high speed test capability. The initial specification of the new tester has begun, and we hope to have a working prototype in about a year.

We have also begun work on a next generation set of tester chips. The goal of this project is to develop pin drive electronics that can set edges to about a 2ns resolution, and operate at a vector rate of 30-40MHz. We have finished the preliminary design of the pin drive electronics and hope to have a test chip out by the middle of next year.

Staff: M. Horowitz, J. Miyamoto, J. Gasbarro

References: [Miyamoto 87]

3.2 Testable CMOS Design

Static CMOS circuits possess certain unique failure modes that cannot be detected by a stuck-at fault test set. Many ATPGs use a switch-level circuit model to accommodate CMOS stuck-open and stuck-on faults. However, the effectiveness of switch-level ATPGs is limited due to their inability to process large circuits. As an alternative, we investigated DFT techniques to solve CMOS testability problems.

For stuck-open faults, a testable circuit structure and its test scheme are presented in [Liu 86]. This circuit structure requires the addition of an inverting buffer to every logic gate that drives

other logic gate(s). Stuck-open faults in this circuit structure can be detected with a simplified 2-pattern test scheme that remains valid under stray delays.

For stuck-on faults, a testable circuit structure and its test scheme are presented in [Liu 87]. This circuit structure consists of specially designed logic gates that have no undetectable stuck-on faults. The test scheme uses a 2-pattern test to detect a stuck-on fault in the circuit structure.

The above two circuit structures can be combined into a fully-testable combinational circuit structure. The two test schemes can be merged into a 3-pattern test scheme. Test patterns for this scheme can be generated by a gate-level ATPG instead of a switch-level ATPG.

Staff: EJ. McCluskey and D. Liu

3.3 Parametric Testing and Diagnostics

During the last six months the testing group completed the investigation of the ion implant monitor structure, as described in Sec. 3.3.1, and concluded processing the first lot of comprehensive, interrelated set of CMOS end-of-process test structures aimed at process problem diagnosis, elemental defect density extraction, and studies of defect clustering and yield prediction. This work is described in more detail in Sec. 3.3.2. Additional accomplishments aimed at establishing a reliable testing facility to support the above activities and to facilitate the transfer of the research findings to industry are summarized in Sec. 3.3.3.

3.3.1 In-process Monitors

Ion implant monitors for dosimetry, channeling, shadowing.

The effect of wafer orientation, angle and tilt on ion channeling in Silicon has been empirically observed to be minimized for particular values of these parameters [Turner 85]. Shadowing is the result of the inadvertent blocking of an implant due to the existence of "tall" features on the wafer or due to high energies where the effect of any surface feature exaggerates the shadowing. A set of ion implant monitor electrical test structures designed to measure dose uniformity, channeling and shadowing effects for the purpose of implanter calibration and evaluation will be integrated onto one wafer. This will enable the simultaneous monitoring of these effects in a single implant.

The mask set has been designed as a generic set such that the same mask set and similar process may be used in performing experiments for p-type and n-type implants, although the substrate and epitaxial layer polarities need to be reversed. No metallization is required. Probing is performed by using standard probes on heavily doped contact regions.

The substrate used in the initial experiments is p-type with a n-type epitaxial layer. Isolation for the structures is achieved by forming mesas using KOH.

The dosimetry structure uses a Van der Pauw configuration in which the central portion contains the light dose implant and the contact pad areas are heavily implanted. This is similar in principle to the structure outlined in [mccarthy 86].

The shadowing structure uses a self-aligned Tee layer arrangement of oxide on poly. The oxide is used to block the implant and provide the shadow. Electrical linewidth measurements are

performed to determine the extent of assymetry in the implanted regions. Sheet resistance information for the layer will already have been obtained using the previous structure.

The channeling structure uses the JFET technique with the implanted dose acting as the control gate. Three different structures are implemented in this process. Structure (a) provides information on the junction depth when the implant impinges directly on the silicon surface. Structure (b) randomizes the incoming implant with a dielectric screen. Structure (c) is the reference structure from which the local epitaxial layer uniformity may be extracted. The monitored implanted is blocked in this structure.

3.3.2 End-of-process Monitors

The end-of-process monitors consist of thirty two 2x2mm modules arranged in an 8x16mm reticle field [yarbrough 86, lukaszek 86]. Each 2x2mm module consists of one or more unique test structures which in the case of defect density monitors are further divided into geometrically ratioed sub-structures. The parametric module contains several tens of small area structures, including individual transistors, linewidth monitors, etc. The choice of the defect density monitors was guided by the desire to electrically determine all bulk, interface, and topography related defect densities of our two layer metal, n-well CMOS process. Two general sets of structures are employed for this purpose. One set of structures divides the composite process structures (e. g. transistors) into interrelated sets of simplest possible sub-structures to obtain "elemental" defect densities associated with these sub-structures. Another set of structures carefully combines elements of the simplest structures into more complex structures to examine the additive relationships between different elemental defect densities as a prelude to yield prediction from elemental defect densities. This approach is culminated in modules consisting of 160x160 µm ring oscillator structures which will be used as a test vehicle for studying defect clustering and as a final check on the validity of the yield prediction relationships.

The salient features of the 32 end-of-process test structure modules are described below.

• 1. Gate Oxide Module #1

This module contains four test structures for evaluating defects in gate oxides. There are two structures, identical in design but different in size, which evaluate the area gate oxides in NMOS and PMOS devices. Two other structures, one for NMOS and one for PMOS, evaluate the gate oxide integrity along field oxide edge. These structures will also be used to evaluate inversion layer-inversion layer isolation. The sizes of these structures represent approximately 85% of the area and edge length that occur in the MIPS processor chip, our most ambitious IC.

• 2. Gate Oxide Module #2

This module contains two test structures, one for PMOS and the other for NMOS, to evaluate the gate oxide integrity along poly edge. These structures will also be used to evaluate source-drain leakage along the channel, and the junction leakage component of source/drain junctions along poly edge. These structures represent approximately 85% of the physical MOSFET edge length in the MIPS processor chip.

• 3. Drain-to-Source Leakage Along Field Oxide Edge Module

This module contains two structures, one NMOS and one PMOS, to monitor drain to

source leakage along the field oxide edge. They are arrays of transistors, of minimum length and width, connected in parallel by common diffusion busses to minimize the influence of contacts on junction leakage. The number of transistors in these arrays represent approximately 25% of the total number of transistors in the MIPS processor chip.

• 4. Metal to N+ Diffused Region Contact Junction Module

This module contains two structures to monitor area and contact induced leakage of diffused N+ junctions. One of the structures has a large number of contacts, while the other structure has a very small number of contacts. The structure with large diffused area and small number of contacts represents approximately 25% of the N+ diffused area that occurs in the MIPS processor chip. The structure with a large number of contacts has 4 connected subsections which have 1, 2, 4, and 8 times the minimum design rule spacing between contacts. This structures represents approximately 20% of the number of contacts (and associated metal area) in the MIPS processor chip. All contact windows are 2 micron X 2 micron. Successive rows of contacts are offset to facilitate cleavage through contacts for SEM examinations.

• 5. Metal to P+ Diffused Region Contact Junction Module

This module is identical to the Metal to N+ Diffused Region Contact Junction Module except that the contacts are to P+ diffused regions.

• 6. Metal to N+ Diffusion Contact Module

This module contains two identical structures for evaluating yield of contacts between first level metal and N+ diffused regions. Both structures consists of three separate contact chains which are tapped at several places. These two structures contain approximately 25% of the number of metal to N+ diffused region contacts in the MIPS processor chip.

• 7. Metal to P+ Diffusion Contact Module

This module is identical to the Metal to N+ Diffusion Contact Module except that the contacts are to P+ diffused regions. The structures in the module contain approximately 31% of the number of metal to P+ diffusion contacts in the MIPS processor chip.

• 8. Metal to Poly Contact Module

This module contains two identical structures for evaluating the yield of contacts between first level metal and polysilicon. Each structure consists of one continuous contact chain that is tapped at several places. These structures contain approximately 25% of the first level metal to poly contacts in the MIPS processor chip.

• 9. Via Strings Module Metal Composite Serpentine Module

This module contains two structures for evaluating the resistance and yield of vias between first and second level metal. The yield monitor structure consists of a tapped via chain. The other structure consists of several individual via chains of identical lengths and varying numbers of vias to determine the influence of vias on chain resistance. This module contains approximately 25% of the number of vias in the MIPS processor chip.

The metal composite serpentine evaluates the metallization integrity of the complete metallization system. This includes metal2, metal1, metal2 to metal1 vias, metal1 to diffusion contacts and metal1 to poly contacts. It will be used primarily for yield prediction exercises.

• 10.11. Parametric Structures Modules Field Isolation Module

This group of structures is used to collect parametric data for process monitoring and device and circuit analysis. The modules include structures for measuring sheet resistance and linewidth of metal1, metal2, poly, N+ diffusion, P+ diffusion and N-well diffusion. Also included are structures for measuring metal1/N+ diffusion, metal1/P+ diffusion, metal1/poly and metal1/metal2 contact resistance. 17 N-channel and 17 P-channel transistors are included for SPICE and TECAP parameter extraction. Field transistors are included for field threshold voltage data and for device to device isolation integrity analysis. Inverters are included for inverter characteristics extraction. A 127 stage ring oscillator with a Schmitt trigger starter circuit is included for extracting gate delay and power consumption data. 3 N-type and 3 P-type capacitance structures are included for measuring bottom junction, sidewall junction and gate overlap capacitances for SPICE circuit simulation. CMOS latch-up characteristics are evaluated as function of critical spacings between diffused regions of typical CMOS inverters.

The field isolation module contains two structures, one NMOS and one PMOS, which evaluate the integrity of the isolation between diffused regions. They can also be used to monitor metal field threshold voltage. These structures represent approximately 25% of the diffused region area, at minimum design rule width, of the MIPS processor chip.

• 12. Metallization Step Coverage and Photolithographic Proximity Effects Structures Module

The step coverage structures are used to evaluate metallization shadowing and step coverage over increasingly difficult topography created by the underlying layers. Each structure isolates a particular topographical situation which may aggravate lithography and/or step coverage.

The photolithographic proximity effects structures are used to detect systematic problems in metal lithography due to reflections which unintentionally expose resist patterns in adjacent valleys. Two types of patterns are used in each of the 3 photolithographic proximity effects monitors. One pattern consists of metal lines running parallel to metal and/or poly lines. The other pattern consists of metal lines running over a metal and/or poly "waffle pattern" which simulates the worst case photolithographic topography that metallization could encounter.

• 13,14. Composite Metallization Array (Comb) Modules

These modules are used to evaluate metallization intralevel and interlevel shorts on structures emulating the cache memory of the MIPS-X chip. The structures contain equivalent amounts of diffusion, poly, metall and metal2 areas and edge lengths that occur in the cache memory. The amounts of interlayer overlap and intralayer spacing are also equivalent to that in the cache memory. These structures will also be used for metalization yield prediction from defect densities data obtained on metalization decomposition structures.

• 15,16. Composite Metallization Array (Serpentine) Modules

These modules are used to evaluate metallization opens and interlevel shorts. The structure has the same unique design as the Composite Metallization Array (Comb) structure, described above. These structures will also be used for metalization yield prediction from defect densities data obtained on metalization decomposition structures.

• 17-24. Metallization Decomposition Structures Modules

These structures are used to evaluate all aspects of the two level metallization system. They differ from modules 11-14 in that they isolate one aspect of the two level metallization at a time. Consequently, there are 8 structures which evaluate interlayer and intralayer isolation, step coverage, and lithography. All of these structures are used strictly for problem decomposition analysis and elemental defect densities extraction and do not attempt to emulate any aspects of the MIPS-X cache memory.

• 25-28. N and P-Channel Transistor Arrays Modules

These modules evaluate gate oxide integrity, source/drain junction integrity, drain to source isolation integrity, and device to device isolation in one composite structure. Therefore, they are a recomposition of all the decomposition test structures described above, and can be used to verify that the results of all the decomposition test structures, taken collectively, are accurate. They will be used primarily for yield prediction exercises.

• 29-32. Ring Oscillator Arrays Modules

The yield data obtained from these arrays will be used to verify the feasibility of yield prediction using component defect density data obtained from the previously described process partitioning test structures. Defect clustering information obtained from these structures will also be used to evaluate the assumptions underlying the commonly accepted, and typically inadequate, yield formulae, and to examine the feasibility of manufacture of daringly large ICs.

Wafers containing modules 1-16 have been processed and are currently under evaluation. Wafers devoted exclusively to modules 1-32 are now in the final stages of processing. We are presently developing the necessary test software.

3.3.3 Supporting Activities

In order to establish a reliable testing facility to support the above activities and to facilitate the transfer of the research findings to industry, the test group has also accomplished the following during this reporting period:

• Collaborative Testing and Evaluation

The test group recently established a collaborative effort of testing and evaluation of the previously described in-process and end-of-proces monitors with colleagues at Hewlett-Packard Labs. The industrial mentor overseeing this activity is Dr. Dirk Bartelink. The intent of this effort is to jointly develop the appropriate test and data analysis software to assess the performance of the above mentioned monitors, and to accelerate their transfer to industry.

• Installation of Parametric Test System

The collaborative testing activity with Hewlett-Packard Labs will be greatly facilitated by HP's donation to CIS of a complete HP4062B parametric test system, including a HP9836CT control computer, a HP7946A 55 Mbyte Disc/Tape Drive, a HP2671G graphics printer, and a HP7475A graphics plotter. A donation by Rucker and Kolls of their 1032 auto- stepping prober completes the system, and duplicates the test installation at Hewlett-Packard Labs. This will assure compatibility of jointly developed software and facilitate the reciprocal transfer of software and data between the two installations.

• Donation of ENHANSYS Data Analysis Software

The testing group has also negotiated a donation of a statistical analysis software package from ENHANSYS, Inc. The ENHANSYS software, also employed at Hewlett-Packard Labs, will be used for statistical analysis and reduction of the large amounts of data obtainable from our end- of-process monitors.

• Donation of PROMETRIX Omnimap Resistivity Mapping System

The testing group has also secured a donation of PROMETRIX Omnimap Model 111 Resistivity Mapping System for the Center for Integrated Systems.

• Donation of RS1 Data Analysis Software

The testing group has received a donation of a statistical analysis software package from Bolt, Beranek and Newman, Inc. (BBN). This software will be used in addition to the ENHANSYS software. Its broad applications base allows it to be very flexible and it will be used for statistical analysis which will supplement the ENHANSYS software capabilities.

3.3.4 Wafer Processing and Testing

Wafer Processing

Several CMOS wafer lots containing modules 1-16 have been processed and preliminary evaluation of wafers from each lot indicate that the design and layout of the structures is correct. One CMOS wafer lot containing modules 1-32 has been processed and preliminary evaluation of the wafers indicate that the design and layout of the structures in modules 17-32 is also correct.

• Test Software Development

Test software for measurement of all test structures has been written and debugged. Additional refinements of the software is in progress.

• Test Hardware Development

Hardware refinements have been completed on the Rucker & Kolls 1032 prober to reduce stray leakage currents to the sub-picoampre range.

• Test Measurement Results

Preliminary testing has been done on wafers from several CMOS lots. Initial analysis show that most of the test structures are producing expected results. However, some test structures revealed problems in our process and have played an

extremely vital role in debugging and correcting these process problems. Additional work is underway to refine data acquisition techniques and to establish appropriate measurement conditions.

• Test Data Analysis

Two statistical software packages have been installed for use in data analysis of the measurement data. At present, we have begun the learning process for using these software packages and exploiting their capabilities. Additional work is needed to determine the appropriate data presentation methods. This work is in progress.

3.3.5 Electrical Alignment Test Structures

Introduction

Semiconductor process characterization is critical for control or for predicting the capabilities of a process. Characterization data can be obtained from test structures. One type of test structure, the alignment test structure, gives the relative alignment between two masks. Alignment information can be used to identify problems with the wafer exposure equipment, or determine layout design rules. This report will present the design of a complete set of alignment structures.

Design Requirements

There are two primary goals in designing this set of test structures. One is to create a complete set of structures. In other words, we wish to obtain all the possible critical dimensions and alignments from test structures. The other goal is to obtain the most accurate set of structures possible. Accuracy is crucial because one application for these structures, determination of design rules, requires maximum accuracy. This is because design rules usually push the equipment close to its limit, allowing little room for uncertainty in measurement.

Two possible types of structures, electrical or visual, exist for measuring alignments or critical dimensions. Only electrical structures will be designed because of their advantage in automating the measurement process and because of the superior accuracy of electrical structures over visual structures. However, it is possible to have several electrical structures which perform the same function, but in different ways. If the same purpose can be served by more than one structure, the most accurate structure must be determined. In some cases, the most accurate structure will be obvious from theoretical analysis. In others, more than one structure must be designed and fabricated, and the most accurate determined empirically.

A single misalignment vector will be obtained from two nearly identical test structures: one structure for the x component of the vector and the other structure rotated ninety degrees for the y component. Fig. 1 shows two pair of alignment structures. The four structures will give two misalignment vectors. Note that each pair of structures are located as close as possible to each other since misalignments may vary across the chip.

One noteable layout restriction applies in designing these test structures. Because the MEBES mask exposure system writes in blocks 128 microns high and 1024 microns wide, there may exist discontinuities in the masks at the boundaries of these areas. Since we are designing for maximum accuracy, these discontinuities must be avoided. Fortunately, the pad dimensions of 160 microns between centers gives four pads the same period as five exposure blocks of 128 microns high. This means that if the pairs of test structures are designed to be four pads high,

and if the structures are placed top to bottom, the horizontal exposure boundaries will lie in the same place for all test structures. This allows us to design around the boundaries without regard to structure placement.

Alignment Structures

The state of the s

200322

The Stanford 2 micron CMOS set of masks has been chosen as the work vehicle in order to understand the requirements that the alignment structures must fulfill. The list of masks, in the order that they are applied, is as follows:

- 1. N-well definition
- 2. Active field defines the nitride areas for forming the thick oxide
- 3. N-well protect this, in addition to the active field mask, defines where the field threshold implant goes.
- 4. N-channel implant this is essentially another n-well protect
- 5. Polysilicon
- 6. N+ source/drain the edges of this mask are the same as the active field minus the P+ s/d regions and bloated 1.5 microns.
- 7. P+ source/drain the edges of this mask are the same as the active field minus the N+ s/d regions and bloated 1.5 microns.
- 8. Contact
- 9. Metall
- 10. Via
- 11. Metal2
- 12. Pads

Note that the N+ s/d and P+ s/d masks simply cover the P+ s/d region and N+ s/d regions, respectively, and that their edges lie over the thick oxide. This implies that their misalignments will not be noticeable unless they are very far out of alignment and cover or expose part of the wrong thin oxide region. Therefore, under good alignment conditions, the N+ s/d and P+ s/d edges on the wafer will both be defined by the thick oxide edges (active field mask).

The electrical structures considered here can be classified into four general types. We call these the split resistor, tapped resistor, transistor width and digital vernier structures. If an alignment can be measured by two types of structure, and if one type is not clearly superior to the other, both will be built and their accuracy compared.

Split Resistor Structure

Split resistor alignment structures can be designed if the two masks whose relative alignment is being measured define the opposite edges of a conducting layer. Figs. 2 through 5 are examples of this type of structure. Each structure gives the horizontal component of misalignment. The structure which gives the vertical component is not shown since it is simply the same structure rotated by 90 degrees. In addition, not all of the pads are shown. One half of each structure gives the misalignment measurement, and the other half gives the sheet resistivity which is used in the misalignment calculation. Both halves also give other useful information such as critical

dimensions which is not needed for the misalignment measurment.

Operation and Measurements Consider the poly to active area alignment structure of Fig. 2 as an illustrative example. The structure works as follows. A constant current, l, is applied between the top two diffusion contacts and four voltages are measured, two in each of the upper and lower halves of the structure. If the alignment is perfect, then the bottom two voltages should be the same, since the poly will divide the two diffusion strips equally. If not, then the poly will divide the diffusion strips unequally and the voltages will give the misalignment through the formula:

$$\Delta W = (I\rho_s L/2)((1/V_1) - (1/V_2))$$

where V_1 is the voltage on the side of the structure towards negative poly to active area misalignment (the left, if +x goes to the right), L is the distance between taps, and ρ_s is the sheet resistivity determined from the top half of the structure.

Sheet resistivity can be found from the top of the structure because we have two measurements and two unknowns. The two unknowns are the sheet resistivity, ρ_s , and the difference between the designed and actual widths of each of the lines, E. For example, if the lines are designed to be 3 and 4 microns and are 3.2 and 4.2 microns after fabrication, then E is .2 microns. The formulas giving these parameters are the following:

$$\rho_{s} = ((W_{3} - W_{4})/IL) ((1/V_{3}) - (1/V_{4}))$$

$$E = ((W_{3} - W_{4})/V_{3}) (1/V_{3}) - (1/V_{4}) - W_{D3}$$

where W_3 and W_4 are the actual widths of the top two segments and W_3 - W_4 is known because it is fixed by the design, regardless of the actual values of W_3 and W_4 . W_{D3} represents the designed width corresponding to W_3 .

Design The structure must be designed to minimize measurement errors. To do this, three sources of error were identified that may be affected through the design. Consider only the bottom half of the structure, and define L as the length between taps, W as the designed distance between the poly edge and thick oxide, or the designed width of each diffusion strip, and D as the tap width. The design will then minimize the measurement error for each source of error listed below as follows:

- 1. Finite tap width the finite width of the tap, D, perturbs the current flow through the long diffusion strip, and decreases the measured resistance: maximize W, L, and minimize D
- 2. Surface and side non-uniformities layer thickness and width variations introduce local disturbances: maximize L
- 3. Machine measurement error the voltmeter itself has a limit to its accuracy, which suggests we should maximize the fractional change in the measurement: minimize W

Of these, we were able to obtain analytical expressions for (1) and (3).

The effect of finite tap width was analyzed in [Hall 67] through the method of conformal transformations. When the formula (48) from [Hall 67] is applied to determine the error in measuring the misalignment, it was found that the errors from each side of the structure will cancel at zero misalignment, and that they increase with increasing misalignment. However,

unless the misalignment is very large, the measurement error remains very small. For a misalignment of .1 microns, L = 60 microns, and D = 3 microns, the error is less than .01 percent of ΔW as long as W is greater than 3 microns. In addition, all the dimensions will scale.

The other primary source of error, that due to the voltmeter limitations, can be considerably greater than the tap width error. The percent error can be easily derived as being $We/\{\Delta W\}$, where e is the percent accuracy of the voltmeter (A typical best accuracy for a digital multimeter is approximately .001%). These results suggest that the design should incorporate the smallest W possible.

In consideration of the three sources of error described above, the following layout guidelines were established:

- 1. L: maximum possible for the horizontal alignment structure, L is limited by the 128 micron exposure window size. 110 microns was chosen for the horizontal alignment structure, and 220 microns was chosen for the vertical alignment structure.
- 2. W: minimum design rule
- 3. D: minimum design rule

The structures which follow these guidelines should give an error of less than .3 percent error for $\Delta W = .01$ due to the three sources of error listed above.

Tapped Resistor Structure

Tapped resistor structures can be designed to measure the alignment between two masks if those masks define two conducting layers: one which is a contact to the other. Figs. 6 through 12 are examples of this type of structure. The structures shown give only the vertical component of misalignment, and not all the pads are shown.

Operation and Measurements The structure operation is described in [Buehler 80]. A current is forced through the length of the structure. Three voltages are then measured, one between the two topmost taps, V_1 , one between the second tap and the contact, V_2 , and one between the contact and the bottom tap, V_3 . The misalignment is then

$$\Delta S = (L/2) (V_3 - V_2)/V_1$$

where Δ S is the misalignment in the vertical direction, L is the distance between the two topmost taps, and positive misalignment of the contact mask to the contacted mask is up.

Design Several issues also must to be confronted in this design in order to minimize measurement error. They are listed below, where a is the width of the contact divided by the width of the conducting strip, and S is the distance from the edge of the contact to the center of the tap. The design will minimize the resistance measurement error for each source of error listed below as follows:

- 1. Finite tap width: maximize S, W, minimize D
- 2. Current patterns due to crowding through the contact (for contacts of high conductivity relative to the conductance of the contacted layer): maximize a, S
- 3. Added resistance due to current crowding: maximize a

- 4. Machine measurement error: minimize S
- 5. Surface and side nonuniformities: maximize S

Eqn. (48) from [Hall 67] can be used for analyzing the error due to the finite tap width. The result is that such errors cancel for all values of D, S, W, and misalignment. Therefore, error due to 1 above is zero.

Similarly, the added resistance due to current crowding cancels when the misalignment is calculated. Therefore, the error due to (3) is also zero.

The current pattern in this structure, assuming (worst case) the contact is of much higher conductivity than the contacted layer, can be calculated from eqns. (37) and (38) of [Hall 67]. The results indicate that for nearly all values of a, the current patterns settle before S = 1.5W. Contact width, a, has little effect on the current pattern once this point has been passed.

Machine measurement error is affected similarly to the split resistor structure. The percent error is approximately $Se/\Delta S$, where e is again the percent error in the multimeter measurement.

The results indicate the following guidelines for design of the tapped resistor structures:

- 1. a: minimum design rule
- 2. W: a + minimum design rule contact surround
- 3. S: 2W
- 4. D: minimum design rule

These guidelines will produce an error due to the machine measurement error of approximately 1 percent in ΔS , assuming a W of 5 microns and a ΔS of .01 microns.

Transistor Width Structures

A transistor width structure can be designed to measure the misalignment between two masks if those masks define the two sides of a transistor. Since the current through a transistor is proportional to the transistor's width, misalignment width differences will be detected as differences in current. Fig. 14 is an example of a pair of structures to give the horizontal (top structure) and vertical (bottom structure) components of misalignment between the poly and active area masks. Figs. 15 and 16 give expanded views of transistors for this structure and for a structure to measure the misalignment between the n-channel implant and active area masks.

References: [Chen 85]

Operation and Measurements Consider the top half of Fig. 14 as an example. An equal voltage is placed on each of the center contacts to the four rectangular regions surrounded by poly (Fig. 15). This is the drain voltage for four transistors. Similarly, a voltage is applied to the common source contacts at the top and bottom of each half structure. These four source contacts are connected to a common pad. Also, a gate voltage which is above the transistor threshold is applied to all four poly gates.

The two transistors on each of the half-structures at the top of Fig. 14 are designed to have identical transistor widths. However, the two transistors in the left half-structure have wider

transistors than the two transistors in the right half-structure. Any two adjacent transistors will have drain currents whose difference will be proportional to the transistors' difference in width, or the mask misalignment. To calibrate the measurement obtained from the two adjacent transistors, and to arrive at a misalignment value, a third transistor is needed. This can be either one of the transistors in the other half-structure. The remaining transistor, the one adjacent to the calibration transistor, is not needed, but allows several readings to be made on different pairs of transistors and with different transistors for calibration. It provides symmetry to the structure and allows checking to be made.

We indicate the half-structure by i, where i=1 is the left half structure and i=2 is the right half structure of Fig. 14. The transistor is represented by j, where j=1 is the left transistor of either half-structure, and j=2 is the right transistor of either half structure. With this, the misalignment can be calculated as follows:

$$\Delta W_{ij} = \Delta W_D(l_{i1}-l_{i2})/(l_{1j}-l_{2j})$$

In this equation, Δ W_{ij} is the misalignment of the poly to the active area when the misalignment is measured in half-structure i and the calibration made with the transistors on side j. Δ W_D is the designed value of the width difference between transistor 1j and 2j. Therefore, four measurements of the misalignment can be made with the same test structure, and all measurements should agree.

Design The tradeoffs in the design are relatively simple: maximize the proportion of current through the parallel sides of the transistors and at the same time, allow for uniform current patterns in both the source and drains. The design that seems to satisfy these criteria is as large a structure as possible in order to give uniform current patterns. In the poly/active area alignment structure, the transistor has minimum length $(2\mu m)$ gates on the parallel sides, and $10\mu m$ on the third side. The width of the parallel sides is limited by the surround area required by the contact inside the rectangular drain and is $7\mu m$ for the narrow transistor and $14\mu m$ for the wide transistor. However, on the n-channel implant/active area structure, the channel is determined by the absence of implant, and therefore its size is independent of the contact area. Its widths are $3\mu m$ for the narrow transistor and $9\mu m$ for the wide transistor. In either case, ΔW_D is $6\mu m$.

Digital Vernier Structures

A digital vernier structure can be designed to measure the misalignment between two masks if those masks define layers which can contact each other. These structures allow misalignment measurements to be made without the need to take the difference between two resistance or current measurements. This property is useful in determining misalignment to metal layers since the resistivity of metals is very low. However, there are several disadvantages to this type of alignment structure. First, its upper bound accuracy is limited to the resolution that can be obtained in offseting one edge relative to another edge on a single mask. To obtain the smallest resolution possible often requires multiple mask exposure, which can be a very cumbersome and imprecise process. In addition, the resolution of the mask exposure is often not the same resolution achieved in the fabricated structure due to random irregularities in the layer edges. Another problem that these structures present is that they give a large number of digital "bits" as data. This requires either a large number of pads, or a multiplexer circuit implemented in working logic gates. Either option may not be possible in some cases. However, even with these limitations it may be a viable test structure for metal alignments, and therefore will be explored. Three structure were designed, and are shown in Figs. 17 through 19. These

structures measure metal 1 to contact, via to metal 1, and metal 2 to via alignments, respectively.

References: [Yamaguchi 86]

Operation and Measurements The structures are designed such that each contact is a different distance from the edge of the metal (this cannot be seen in the figure for reasons that will be explained in what follows). Some contact edges are designed to lie on top of the metal, and some are designed not to touch the metal at all. To measure the misalignment, each contact is tested for continuity to the metal. Since the structure contains a similar set of contacts on each side, the measurement is independent of layer widths. The misalignment is given by

$$\Delta W = (1/2)(C_L - C_R) SP$$

where C_L and C_R are the number of high conductivity connections to metal on the left and right, respectively, and SP is the spacing between contact edges, assuming equal spacing.

Design and Implementation The design of the structure is very straightforward. On each side, the center contact is placed such that its inside edge corresponds to the edge of the metal layer. The contacts toward one end are placed such that each edge is offset from that of the center contact in the amounts .1, .2, .3, and .5 microns, respectively as the contacts become more distant from the center. The contacts on the other side of the center will be offset by -.1, -.2, -.3, and -.5 microns.

Because the mask making facilities available are limited to a dot size of .25 microns, special techniques must be employed to obtain smaller offsets. Mask exposure is to be done in nine steps, and nine different types of contacts must be specified. The first type of contact is the standard contact and is exposed along with all other contacts on the mask. The second type is to be exposed after the first, and the mask translated in the x direction by .1 microns. For the third type, the mask is translated in the x direction by .2 microns, and so on through the fifth mask type. For the sixth mask type, the mask is translated in the y direction by .1 microns, and so on. This is the reason that the offset between contacts is not shown in Figs. 17 trough 19: the contacts have been defined but not translated appropriately.

Complete Set of Alignment Structures

The structures shown in Figs. 2 through 12 constitute a complete set of alignment structures for the Stanford 2 micron CMOS process. By a direct measurement, or through several measurements and some vector addition, the relative alignment of one mask to any other can be determined. The structures in Figs. 14 through 19 are redundant with these in what they measure, but it may be found that they are more accurate. All of these designs are included in the following pages.

Future Work

This work is part of a larger project on determination of design rules. Work planned in this project includes design of critical dimension and defect density test structures, and system design to translate the data obtained from the test structures into design rules.

Staff: Dr. Terry Walker, Dr. W. Lukaszek, Anthony McCarthy, Willie Yarbrough, Greg Freeman, Tsu-Chang Lee

References: [mccarthy 86, yarbrough 86, lukaszek 86]

4 Fast Turn-Around Laboratory

4.1 Fable: Knowledge Representation for Semiconductor Manufacturing

The Fable project has made significant progress over the past months. We learned important lessons from our first efforts in this area, and are forging ahead with renewed energy and excitement.

Our recent accomplishments are as follows:

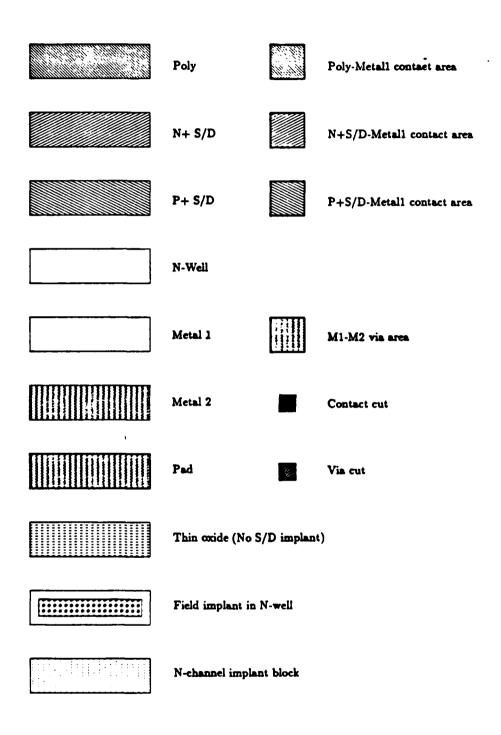
- 1. We have identified the important lessons learned from our initial designs and implementations of Fable.
- 2. We have broadened Fable to capture more of the knowledge necessary to describe and carry out semiconductor manufacturing.
- 3. We have identified and acquired a very powerful set of hardware and software tools to support the continued development of Fable.
- 4. We have identified and started four applications projects that will serve both to determine specific requirements for Fable and to implement specific parts of Fable to satisfy these requirements.
- 5. We are conducting, for the first time, a Stanford course in "Automation of Semiconductor Manufacturing" (CS412/EE391).
- 6. We have started a nationwide electronic mailing list, IC-CIM, for discussion of topics in computer-integrated manufacturing of integrated circuits.

4.1.1 Lessons

The principal lessons learned from Fable's first three years were the following:

- 1. The Fable problem is largely a problem in representing, acquiring, and using a broad variety and large amount of knowledge about semiconductor manufacturing. The procedural model initially assumed by Fable was not adequate to express the knowledge required. We are now using a more powerful and more general model -- knowledge representation.
- 2. The Fable problem is more difficult than we anticipated. Its solution requires state-of-the-art tools in symbolic computing, knowledge representation and inference. We have acquired and are now using such tools, including Al workstations (TI Explorers) and knowledge engineering tools (KEE and SimKit).
- 3. The Fable problem requires more collaboration than we anticipated. Its solution requires close interaction between experts in IC processing, computer science, and manufacturing, and equally close interaction between university and industry. We are working hard to establish both kinds of collaboration.

Lange parameter personal presentation of presentations of presentations of the presentation of the present of



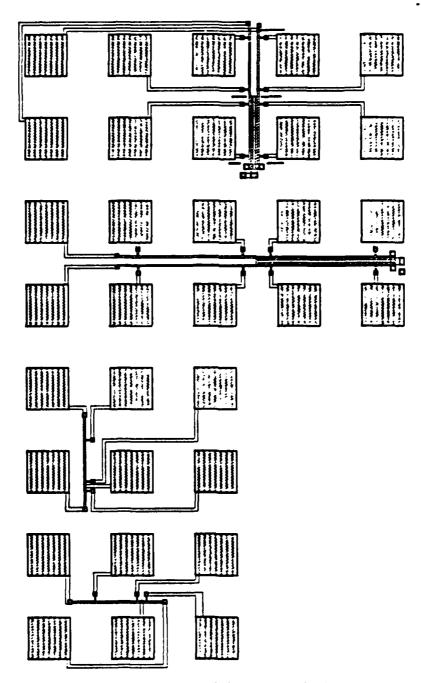


Fig. 1: Four electrical alignment test structures

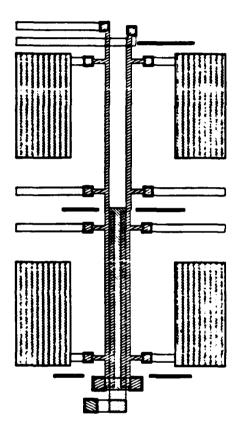


Fig. 2: Poly to Active area

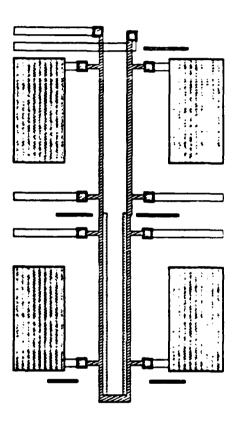


Fig. 3: N+S/D to Active area

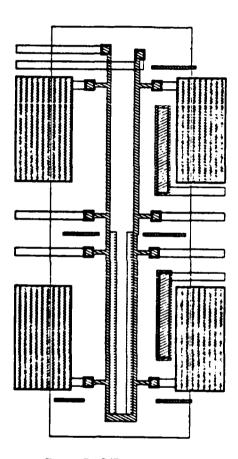


Fig. 4: P+S/D to Active area

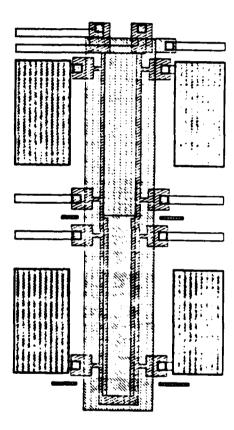


Fig. 5: N-channel implant to Active area

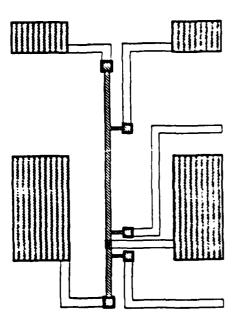


Fig. 6: Contact to Poly

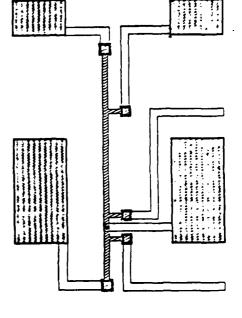


Fig. 7: Contact to Active area

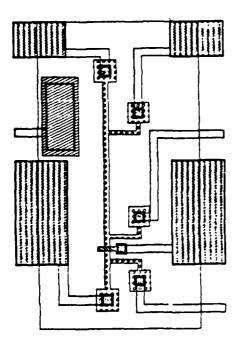


Fig. 8: N-well protect to Active area

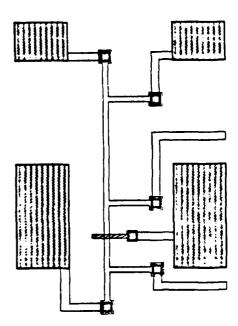


Fig. 9: Active area to N-well

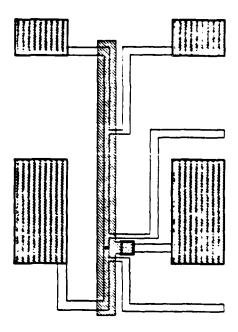


Fig. 10: Metal 1 to Contact

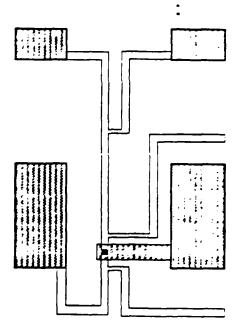


Fig. 11: Via to Metal 1

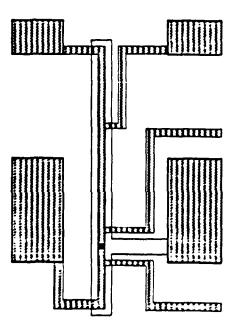
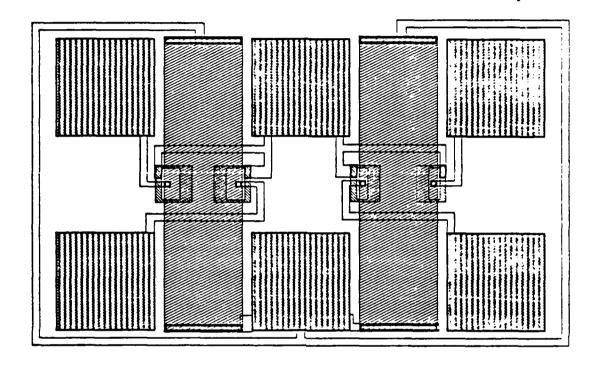


Fig. 12: Metal 2 to Via



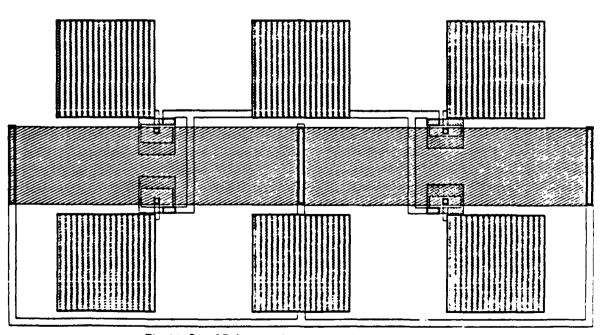


Fig. 14: Set of Poly to Active area transistor width structures

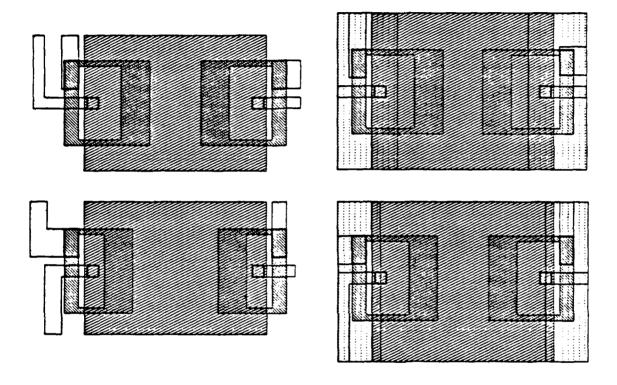


Fig. 15: Poly to Active area

Fig. 16: N-channel implant to Active area

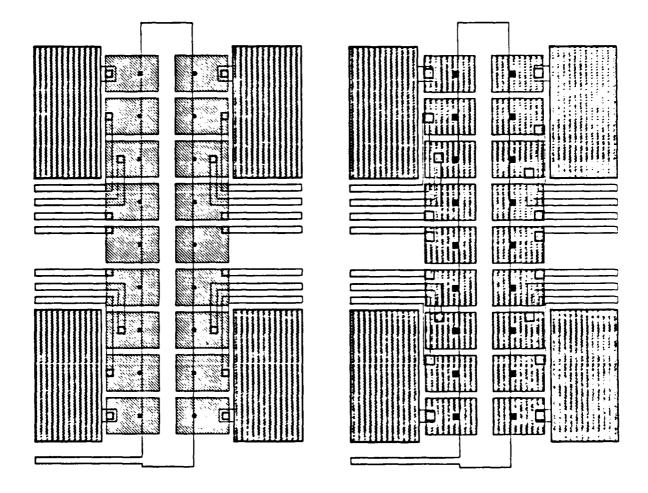


Fig. 17: Metal 1 to Contact

Fig. 18: Via to Metal 1

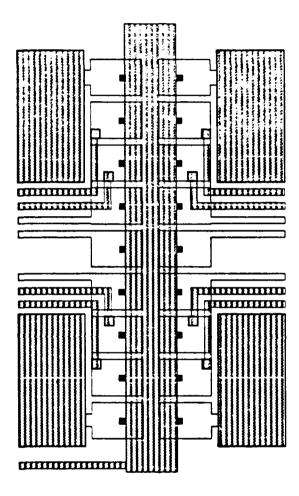


Fig. 19: Metal 2 to Via

4.1.2 Knowledge and Representation

Much of the knowledge required by an automated fabrication system cannot be effectively encoded as procedures. To reason about a fabrication process, for example, requires information not just of the process, but also information about the process. An example of knowledge that is difficult to express procedurally is the following:

Recipe CM142 is very similar to Recipe CM101, which was begun 24 hours earlier. If the initial parametric data from CM101 are marginal or worse, it is recommended that CM142 be suspended until the problems are identified and corrected.

Although it is possible to encode such information procedurally, anore declarative representations of such information can be easier to understand and modify.

The automated fabrication line needs to know about more than just the processes that run on it. To execute the processes, the fab line needs to have knowledge about equipment, materials, schedules, and preferences, for example. We will need to use a broad variety of knowledge representation mechanisms to effectively capture knowledge about such a broad variety of entities.

4.1.3 Knowledge Tools

To efficiently develop prototypes of systems to gather and use such knowledge for automation, we need powerful tools. Thanks to Texas Instruments and IntelliCorp, we now have access to a very powerful software development environment, incorporating the TI Explorer workstation and the KEE knowledge engineering environment. In one project, this environment has enabled us to do, in one month, more than could have been accomplished in six months in a more conventional Unix environment. We expect this productivity to extend to the entire Fable project.

4.1.4 Applications Projects

CIS researchers have identified four new projects related to automation of semiconductor manufacturing, and have begun to work on them.

The first such project involves semiconductor factory simulation. A group of three students have implemented a queueing model simulation of a semiconductor fabrication line, and are investigating algorithms for scheduling the simulated line. The students are developing a process specification language (a prototype of the procedural component of Fable) to permit the simulated line to run realistic fabrication recipes.

The second project is focused on developing intelligent processing equipment. The general idea is to closely couple a piece of semiconductor manufacturing equipment, such as an ion implanter, with a state-of-the-art Al workstation. The linkage will give the workstation direct access to the sensors and controls of the processing equipment. The linkage permits and encourages the development of knowledge-based software to support:

- 1. interactive monitoring and control of the processing equipment through a graphical interface;
- 2. high-level communication between host computer and the augmented processing equipment;

- 3. automated diagnosis of equipment and process problems;
- 4. automated monitoring and control of the recipe running on the processing equipment; and
- 5. design and simulation of process steps to run on the equipment.

Part of this second project involves developing languages (as part of Fable); for describing recipes specific to particular classes of processing equipment, such as ion implanters.

A third project is investigating expert systems for automatically interpreting electrical measurements from test structures. When the electrical measurements deviate significantly from those predicted by circuit simulation (using SPICE, for example), we would like an expert system to tell us which physical parameter is likely to be responsible for the deviation, and the amount of the deviation.

A fourth new project is implementing the SECS-I and SECS-II protocols for communicating between host computers and semiconductor processing equipment.

4.1.5 Object Oriented SECS-I/II Equipment Communications

The main project started at this time was the development of a SECS-I and SECS-II capability under Unix 4.3 using the IP networking protocols with the Berkeley Unix socket mechanism. This will eventually allow any computer system on our ethernet network to access a piece of process equipment equiped with a SECS interface. A master's level student has started work in this area. The adoption of the Berkeley laboratory control system also continued. This system will provide an initial test enviorment for a number of automation experiments. The initial machine being used for testing the work is a Varian 350D implanter equiped with a very complete SECS-I and II interface. Information has been transferred to and from the implanter and a DEC VAX-11/750 but only in a rudimentary manner. The programming language C++ is also being considered for developing the control programs on the host computer. C++ is an object oriented form of the C language and provides a good means of encapsulating the various levels of detail needed in such an interface. It has the added advantage of actually being compiled into ordinary C language used on all Unix systems and is therefore very portable.

Another student working on a class project has begun developing an object oriented frontend to work with the SECS-I interface program. The SECS-I program (known as a "daemon" in Unix terms) will provide a generalized interface between different computer systems on our ethernet network and a number of pieces of process equipment. The object oriented SECS-II programs written in C++ will be able to access this equipment from any main computer (VAX-11/750 or 11/780) or workstation (SUN, DEC MicroVax, or TI Explorer). This allows any machine with the correct capability (AI (Lisp), Smalltalk, or C++) to establish a direct link with the process equipment and perform experiments directly with the machine.

4.1.6 Collaboration

Automating semiconductor manufacturing will require expertise from a number of fields, including semiconductor processing, computer science, and IC manufacturing. The Fable project welcomes the participation of people from all these fields.

Automating semiconductor manufacturing will require collaboration between researchers and

implementors in both university and industry. The Fable project welcomes the participation of industrial colleagues and encourages students to obtain first-hand experience with industrial semiconductor fabrication lines.

4.1.7 New Automation Course

"Automation of Semiconductor Manufacturing" (CS412/EE391) is being offered for the first time this fall at Stanford. The course includes lectures on semiconductor manufacturing, existing automation in the semiconductor industry, object-oriented databases, scheduling, knowledge representation and expert systems for semiconductor manufacturing, automated interpretation of test structures, and other AI topics. To learn from our industrial colleagues and to help them learn about topics in semiconductor automation, we have welcomed them to the course as both speakers and listeners.

The announcement of the course is included below.

4.1.8 New Electronic Discussion Group

To encourage and facilitate discussion of topics in computer-integrated manufacturing of integrated circuits, we have started an electronic discussion group called IC-CIM. An announcement of the mailing list is included below.

COURSE ANNOUNCEMENT: CS 412/EE 391

Subject: Automation of Semiconductor Manufacturing

Time: Th 11:00-12:15 (first meeting October 2):

Location: CIS 101 (Center for Integrated Systems, Stanford)

Credit: 1-3 units (1 for attending lectures, 3 for project)

Purpose: To explore and exploit opportunities for automation

of semiconductor manufacturing processes.

Topics: State of the art of semiconductor manufacturing automation in the lab and in industry. Increasing the effective use of computation in the design and control of manufacturing processes, through the combination of AI, computer graphics, and simulation. Designing the intelligent, interactive factory.

Format: An explicit goal of the course is to encourage cross-fertilization and collaboration between the fields of AI and semiconductor manufacturing. The course will include discussion of selected papers from both fields and participation in interdisciplinary team projects. Potential project topics include intelligent interactive interfaces for semiconductor processing equipment and the use of AI in design, simulation, monitoring, control, and diagnosis of semiconductor manufacturing processes.

Prerequisites: Consent of instructor. A familiarity with AI (at the level of CS 223) or IC fabrication (at the level of EE 312) is recommended.

For further information contact:

Jay M. Tenenbaum, Consulting Professor, Computer Science (415) 496-4699 or Tenenbaum@SRI-KL.ARPA

Byron Davies, CIS Industrial Visitor (415) 725-3714 or Davies@Sierra.Stanford.EDU.

IC-CIM: A New Mailing List for Discussion of Computer-Integrated Manufacturing of Integrated Circuits

To join: Send your net address to 1C-CIM-Request@Sierra.Stanford.EDU.

IC-CIM is a new electronic mailing list for discussion of Computer-Integrated Manufacturing of Integrated Circuits. IC-CIM is maintained and moderated at Stanford University, within the Center for Integrated Systems.

The list of addresses for IC-CIM was initialized from lists supplied by Dave Hodges at Berkeley, Andrzej Strojwas at CMU, Paul Penfield at MIT, and Byron Davies at Stanford. A number of industrial researchers have also been added to the list. New participants are welcomed from both university and industry.

IC-CIM is a moderated mailing list. When messages are sent to IC-CIM, they are gathered into digests of three or four messages and sent out this way to the mailing list. Readers will not be bothered by misdirected, inappropriate, or too frequent messages.

IC-CIM is open to discussion of any topic related to computer-integrated manufacturing for integrated circuits. Examples of IC-CIM topics are:

System architectures: computer hardware, software, networks Al and expert systems for semiconductor manufacturing Processing equipment: capabilities, user interfaces Fabrication simulation
Scheduling and optimization
Process and equipment modeling
Process specification and design systems
Manufacturing databases
Data and knowledge representations for:

- o the semiconductor manufacturing line
- o fabrication equipment
- o wafers at each stage of fabrication
- o other fabrication materials (e.g., gases)

Training aids for operators, supervisors, engineers

Integration of manufacturing, design and test systems

Staff: B. Davies, M. Tenenbaum, P. Asente, L. Adams, E. Kirshenbaum, E. Wood.

Related Efforts: Hodges and Katz (Berkeley), McIlrath (MIT).

4.2 Microlithography

Work has been continuing in microlithography in the areas of e-beam mask making and direct write, optical and thin-film lithography, and inspection algorithms using the template matching scheme. Support work on Mebes and the Ultratech stepper for fast-turnaround and other runs has been carried out.

4.2.1 Electron Beam Lithography

In this time period, 51 mask sets were generated on MEBES. In addition to CIS users, requests were from the physics and applied physics departments, as well as from the programs of Professors Pease, Swanson, Harris, Gibbons, White, and Angell in the EE department.

In addition to the contact resistance test chips fabricated previously with direct write, another group of contact resistance wafers has been fabricated along with a series of pMOS devices. Both of these sets of devices is currently being tested. Level-to-level misalignment of wafers has been found to be consistant within a lot, process dependent, and which slowly drift with time. To correct for the misalignment, a group of test wafers will be included in each run to measure misalignment. The coordinates of the alignment marks given to the MEBES will then be offset by this value as the correction.

Wafers were exposed and developed with $0.5\mu m$ lines and spaces for the task group on interconnections and contacts. These are to be used by for selective W deposition in $0.5\mu m$ trenches to form both vias and interconnections.

The 1/8 µm contract with Perkin-Elmer EBT has been completed. 5" mask metrology procedures developed under this program were used to evaluate the MEBES machine. Address sizes used were 0.5, 0.25, and 0.1µm with the MEBES calibrated at 0.25µm. The results are:

ı	<u>Goal</u>	Initial Evaluation	Final Evaluation
Line edge roughness	0.05μm	X: 0.045 Y: -	0.039 0.029
Stripe butting	0.10µm	X: 0.015+0.058 Y: 0.3+0.067 (mean + 3sigma)	0.003+0.049 0.3+0.054
Overlay accuracy	0.13µm	X: 0.10 Y: 0.05	not done
Write scan linearity	0.075µm	0.052	0.049

The MEBES performed well in the initial evaluation. All goals were surpassed except in the case of 0.10µm address stripe butting in Y. There we had a 0.3µm mean shortfall. The final evaluation shows a modest improvement in most areas except this 0.3µm Y stripe butting error. This probably indicated a non-linearity in the write-scan length vs. scale dac values. Quadratic-fit write-scan coefficients, where the dac value (DAC) is given in terms of scan length desired (L) by:

$$DAC = C_0 + C_1L + C_2L^2$$

were installed on MEBES to deal with the problem. The program used to calculate the coefficients from metrology data needs further work to work correctly. Metrology procedures developed in this joint development effort have been transferred through Perkin-Elmer to other MEBES sites in the U.S.

4.2.2 Defect Inspection

The first fabrication run for the random defect detection circuit based on the template-set inspection scheme has been completed in our laboratory. It includes a 3.2K maskable content addressable memory array and uses a 2 μ m double-metal CMOS technology. In anticipation of the laboratory shutdown starting in October, we also submitted this design to MOSIS and expect to receive prototype chips in November. Work in the immediate future will include testing of these chips. A prototype inspection system for demonstration will also be built if functional chips can be obtained.

As discussed in previous reports, our template-set approach was initially applied only to the detection of defects. In conjunction with this work, we are also developing a new local chain-coding method to represent key topological properties of the local pattern. Detailed defect classification can then be achieved easily based on such information.

Alternatively, we can extend the template set technique to perform defect detection and classification simultaneously. In this scheme, all templates representing the characteristics of each type of pattern defects (e.g. pinholes, protrusions, and width errors) are collected to form a template set. The local window pattern is then compared to multiple template sets (representing various defect categories of interest) in parallel, thereby providing information on the defect type in addition to location. Our initial study showed that the classification capability can be added in this way without substantial increase in complexity, resulting in a system which allows real-time inspection on standard video images.

4.2.3 Langmuir-Blodgett Films

With an aim to use the Langmuir-Blodgett (LB) films as resist materials for microlithography, experiments have been designed to see if greater sensitivity can be obtained in a polymerizable LB film by doping it with heavy metal ions (such as Cd), which are expected to scatter and absorb more radiation. Specular reflection grazing angle Fourier transform infrared spectroscopy has been used to characterize the degree of polymerization of films, which can be directly related to the dissolution rate in the developing process. Such measurements have been performed on brassidic acid at various electron beam energies and currents, and at different temperatures. The results exhibit an exponential relationship between the degree of polymerization and the exposure dose. Future plans are to characterize the polymerization of cadmium brassidate and compare it with that of the acid, and to expose these films with X-ray radiation.

Staff: R.F.W. Pease, D.H. Dameron, C.C. Fu

References: [Chae 86]

4.3 Processes, Devices, and Circuits

4.3.1 Interfacial Charges and Time-Dependent Breakdown in MOS Devices with Rapidly Grown Ultrathin SiO₂ Gate Insulators

Rapid thermal processing (RTP) is an emerging technology with many important applications in silicon integrated circuits and compound semiconductors. In the past few years, we have been investigating some novel applications of RTP including thin dielectric growth by rapid thermal oxidation and nitridation processes and reactive-ambient annealing of refractory metals. The results have been very promising and we will continue these studies with the goal of placing further emphasis on manufacturability of the RTP applications. The main objectives of this work are to identify and analyze the RTP equipment parameters and manufacturing requirements through its conventional and novel applications and by employing an advanced RTP-based submicron CMOS technology as a technology vehicle. Because RTP is a multipurpose technique (annealing, growth, and deposition processes), the equipment and process models are application-dependent. As a result, a comprehensive physical understanding of these process applications is necessary for developing appropriate application-specific equipment models.

Rapid thermal processing of silicon in oxygen and ammonia ambients is an attractive technique for growth of thin dielectrics such as silicon nitride, silicon dioxide, nitrided oxides, oxidized nitrides, and application-specific dielectrics (such as oxides with a buried layer of nitride near the interface). Multicycle rapid thermal growth processes are suitable for dielectric engineering and in-situ formation of thin layered insulators with a variety of controllable oxygen and nitrogen compositional depth profiles by appropriate design of the temperature and ambient gas cycles.

Nitroxide films are prepared by rapid thermal nitridation (RTN) of SiO_2 films, usually grown in a furnace. Because the preparation of RTN nitroxide requires the growth of SiO_2 before RTN, the oxidized silicon wafers should be transferred from the furnace into the RTN chamber following oxidation. In situ multiprocessing can enhance yield and reduce contamination and, as a result, it is advantageous to perform both the oxidation and nitridation processes via RTP which will greatly simplify the growth of RTN nitroxide. Conventionally, SiO_2 films have been grown in standard furnaces where oxidations are long ($t \ge 20$ min), and lower oxidant partial pressures are required to grow very thin films with good electrical characteristics.

To develop thin layers of silicon nitroxide by rapid thermal nitridation, silicon was first oxidized in a furnace to grow a thin (100 Angstroms) layer of SiO₂: subsequently, the oxidized silicon wasfers were subjected to ammonia at high temperatures (900°C to 1200°C) to convert SiO₂ into layered silicon nitroxide. One motivation for promoting the rapid-thermal-oxidation (RTO) technique described in this section was the growth of initial SiO₂ by RTO instead of furnace oxidation prior to the RTN cycle. The necessary two-step processing in two different pieces of equipment would be eliminated which would greatly improve quality, yield, and process simplicity because RTN can then immediately follow RTO in the same RTP chamber to grow high-quality silicon-nitroxide films on silicon. Extensive new results obtained from the RTO process are reported here, including the initial regime of thermal oxidation of (100) silicon in dry O₂ and the SiO₂ films grown via this technique under a wide range of experimental growth conditions. Investigation of thin SiO₂ growth kinetics in the very short oxidation regime with standard oxidation furnaces is not precise because the transient times involved in furnace processing are much longer than the very short oxidations to be studied; however, RTP has

become a nearly ideal tool for these applications.

The kinetics of rapid thermal oxidation (RTO) of Si have been previously studied. The preliminary electrical data for oxides grown by the RTO process have indicated 13 MV/cm breakdown fields with well-behaved conduction and C-V characteristics. This report summarizes additional results on the ramped-voltage and time-dependent dielectric breakdown (TDDB) characteristics of MOS devices with rapidly grown unannealed gate oxides and the effects of preoxidation cleaning and native oxides on charge-to-breakdown, fixed oxide charges, and surface states.

Rapid thermal oxidation (RTO) of silicon is an attractive technique for growth of thin gate and tunnel dielectrics for submicron MOS technology.

This report presents a summary of our recent research results on the fixed charges (Q_f) , surface states (D_{it}) , and dielectric breakdown of gate oxides grown in a lamp-heated system and their dependencies on the RTO conditions, postoxidation anneal (POA), forming gas anneal (FGA), and preoxidation wafer cleaning. As part of these studies, metal-oxide-semiconductor devices fabricated with tungsten/n⁺ polysilicon composite gates and subhundred-angstrom SiO_2 gate insulators grown by rapid thermal oxidation were characterized by various electrical measurements. The as-fabricated devices with unannealed rapidly grown oxides exhibited breakdown characteristics superior to furnace-grown oxides as evidenced by their excellent breakdown uniformity, an average breakdown field of 15 MV/cm, and an average breakdown charge density of over 50 C/cm² at a stress current density of 1 A/cm². The preoxidation surface cleaning procedure was observed to affect the charge-to-breakdown and the densities of fixed oxide charges and surface states in these MOS structures.

The RTOs were performed on n-type (100) Si wafers at 950°C -1150°C to grow oxides on the order of 80 Angstroms. Prior to RTOs, one group of wafers was cleaned using a modified RCA technique leaving a chemically grown native oxide of 15 Angstroms; the other group of wafers was cleaned similarly but with a final 50:1 H₂O:HF dip and Dl H₂O rinse to remove the chemical native oxide. Gates of MOS devices were a composite structure of n⁺-polysilicon/CVD tungsten. To study the intrinsic properties of rapidly grown oxides and also separate and determine the other processing effects, the characterizations were performed on various splits of wafers: one split without POA and final FGA, one split with POA (at 1038°C for 45 sec in Ar performed after definition of polysilicon gates), and another split with both POA and FGA. The emphasis in these studies was placed on the characteristics of unannealed devices.

The TDDB and trapping phenomena were investigated by the constant-current stress technique. The charge-to-breakdown (Q_{bd}) vs RTO temperature consistently indicates a higher Q_{bd} when the chemical oxide is removed by a final preoxidation HF dip. The increase in constant-current voltage is an indication of net electron trapping which was observed to be much less in rapidly grown oxides compared to furnace grown oxides. For example, in devices with oxide grown at 1050°C the total rise in constant-current voltage up to the onset of destructive breakdown was less than 0.7 V and the average Q_{bd} is over 50 C/cm² at 1 A/cm² stress current density, which is larger than 20 C/cm² measured for furnace-grown oxide under similar stress conditions. Q_{bd} is significantly increased (over 100 C/cm²) at lower stress current densities because of reduced oxide electric field. Q_{bd} was observed to decrease with an increase in the gate area as a result of

the defects present in oxides and dominance of extrinsic defect-related phenomena in the large area devices.

The measured conduction characteristics indicate a Fowler-Nordheim mechanism over more than seven decades of current. The calculated barrier height and effective electron mass in these oxides were found to be $3.29 \, \text{eV}$ and $0.37 \, \text{m}_{\odot}$, respectively ($\, \text{m}_{\odot} = \text{electron rest mass}$). The ramped-voltage breakdown of oxides grown at various temperatures were also investigated. The multiple breakdown measurements on devices with 79 Angstroms SiO_2 indicated an average destructive E_{bd} of 15 MV/cm with excellent breakdown uniformity. The effect of preoxidation surface cleaning on E_{bd} and F-N conduction in devices with 105 and 98 Angstroms oxides grown at the same RTO temperature was investigated. In both cases the average E_{bd} is 15 MV/cm and the breakdown field and its integrity appear to be independent of the preoxidation cleaning; however, the F-N conduction distribution is tighter when the chemical oxide is etched away prior to RTO, possibly because of the slight nonuniformities associated with the chemically grown native oxide.

The C-V characteristics of the unannealed devices indicated distortions because of surface states which could be annealed out by proper POA and also FGA. The C-V and D_{it} characteristics of all devices with annealed or unannealed oxides grown at various RTO temperatures were similar in terms of general behavior; however, the key parameters such as flatband voltage (V_{fb}), and minimum and midgap D_{it} 's were dependent on the preoxidation cleaning and RTO growth conditions.

The dependence of V_{fb} on RTO temperature was determined. In the devices with final HF dip V_{fb} became more negative with higher RTO temperature because of a larger Q_f as a result of a faster oxide growth rate. The devices without the final preoxidation HF dip did not exhibit a similar trend and their V_{fb} values were nearly independent of the RTO temperature. The more positive V_{fb} in the latter indicates a smaller Q_f at the interface when no HF dip is used. The effect of preoxidation surface cleaning on V_{fb} becomes less significant at lower RTO temperatures where both cleaning procedures result in near-zero V_{fb} 's.

The minimum D_{it} 's with the two different preoxidation cleaning procedures were determined as a function of oxidation temperature. In the wafers with the final HF dip, $D_{it,min}$ decreases as the RTO temperature is increased but it appears to be independent of the oxidation temperature when no final HF dip is performed prior to the RTOs. The effect of preoxidation surface cleaning on $D_{it,min}$ is negligible in the high growth temperature regime where surface-state densities are minimum and converge for two cleaning procedures.

Additional studies regarding electrical performance, hot-carrier degradations, and surface mobilities in MOSFETs with rapidly grown oxide gate insulators are in progress.

4.3.2 Dry Etching

Work in this area has continued to be divided between supporting the FTAL CMOS runs and research into understanding and controlling dry etch processes. The support function has focused on improving the reliability of our dual level metal process. While the research efforts have focused on developing diagnostic tools for monitoring and modeling dry etch processes, on studying sidewall inhibitor layers which are believed to be responsible for the anisotropic

properties of a number of etch processes.

Dual Level Metal

Our dual level metal process, which consists of two Al layers separated by a planarized layer of deposited SiO₂, has suffered from two reliability problems which were: 1. Shorts between the metal layers above poly-Si regions. And 2. opens in the top metal layer crossing 1st level metal. The short problem was solved by first going a two step oxide deposition process where the 1st 1000 A is deposited at 300 C and the rest is deposited at 380 C. The initial oxide is used cap the Al and help reduce hillock formation. The second step in eliminating hillocks was to switch from a 1st level metallurgy of composite Al/Si/Ti to a 500 A Ti cap on 5500 A of Al/1%Si. The Ti cap was dramaticly more effective in suppressing hillocks than Ti mixed in the Al/Si.

The open problem in the 2nd level metal was found to be related to shallow trenches left from the planarization of the interlayer oxide at the edges of the 1st level metal. In wet etching of the 2nd metal layer, the etchant was able to undercut resist covered lines by following these trenches. This problem was eliminated by reducing the phosphorus doping at the top of the oxide layer, and by going to dry etching of the 2nd metal layer.

In-situ Monitoring of Plasma Parameters During Dry Etching

The understanding and control of dry etch processes has been held up by the lack of knowledge of the internal plasma parameters. Most etch systems only control power, pressure and gas flow, and do little to directly monitor the plasma discharge set up above the wafer. To improve this situation we have investigated the use of external voltages and currents to monitor the internal plasma parameters. Three parameters of particular interest are the electron density (n_e which controls the generation of the active species, the ion current density (J_i which determines ion flux onto the wafer, and the sheath thicknesses (t_{s1} and t_{s2}) which along with the sheath voltage controls the energy of the ions striking the wafer.

To make use of external electrical measurements from a etch system, we have analyzed the different mechanisms responsible for current transport between the electrodes and have developed a circuit model whose components have been derived in terms of the internal plasma parameters. The circuit components is this model include a resistor for the bulk collision limited electron current, a non-linear resistor for the space charge limited ion current in the sheaths, capacitors for the low energy electron induced displacement current across the sheaths, and a exponential voltage dependent current source for the high energy electron space charge limited current across the sheaths.

This circuit model approach to measuring the internal plasma parameters was tested by using it on a SF_6/O_2 discharge in a parallel plate etch system operating in the plasma mode. Using this model values were obtained for the n_e , J_i , t_{s2} , and t_{s2} . The values obtained for n_e agreed well with both Langmuir probe measurements and equilibrium Boltzman calculations. The values for the sheath thicknesses agreed with optical measurements. The advantage of this method is that this technique is non-invasive and can be incorporated into a plasma etcher with a minimum of hardware.

Sidewall Layers in Dry Etching

During plasma etching, several different sub-processes occur simultaneously and in an

interactive manner. These sub-processes are: (1) chemical etching (2) ion bombardment of the surfaces exposed to the discharge (3) residue formation on the substrate To better understand the overall process, these sub-processes need to be separated and individually studied. Understanding their individual effects and their modes of interaction will lead to realistic models of plasma etching.

Residue formation on the substrate during plasma etching is an important issue in optimizing the process. Residues form as a result of polymerization reactions occurring between the free radicals created in the discharge. They may be a nuisance, affecting subsequent processing steps such as deposition and requiring additional steps for their removal. Further, they may reduce the etch rate. However, they can play an important role in obtaining anisotropic etch profiles; especially in single wafer etchers where higher pressures and lower ion energies are used. In these cases anisotropy is a result of the interaction between residue formation and ion bombardment. Residues form on the sidewalls as well as the floor of the etch profile. Ion bombardment, however, is restricted only to the floor since ions are incident normally onto the surface. This results in thinner residue layers forming on the floor as compared to the sidewalls. The etch rate on the sidewalls is therefore lower than on the floor, due to the greater inhibiting effect of the residues on the former. The result: an anisotropic etch profile. Residues also help in getting better selectivities. For instance, during oxide etching, the presence of oxygen in the oxide decreases the concentration of the residue forming precursors. Once the oxide is etched through and the underlying Si exposed, there is no longer a source of oxygen. Residues form rapidly and the etch rate of Si plummets, leading to higher oxide to Si selectivity. Thus we see that residues have both a beneficial and deleterious effect.

An effort is underway to study the nature of these residues and the effect of ion bombardment on them. This is accomplished by stopping, or diminishing the energy and flux of, the impinging ions. This enables the study of the effect of ion bombardment on the nature of the residue and at the same time, allows sufficient residue to be accumulated for surface analysis to be possible. Ion bombardment is reduced by what we refer to as the "grid" technique. The rest of this report will deal with a description of this technique and some of the preliminary results obtained.

The grid technique involves placing the substrate, which lies on the grounded electrode, underneath a grounded aluminum plate which has a number of 8 mm diameter holes in it. Aluminum grids were placed over half of these holes so that etching occurred only under the open or the grid holes. There is no field between the grid and the electrode since they are both at the same potential. Ions are accelerated across the sheath between the discharge and the grid, but once they penetrate the grid, there is no field to accelerate them further. The distance between the grid and the substrate is many mean-free-paths (of the ion) and hence the ion suffers several collisions before it reaches the substrate. This causes it to lose the energy it had gained during its transit through the sheath, and if it eventually hits the surface, it does so with only its thermal energy.

Residue was collected on pieces of bare Si placed under the open and grid covered holes. The etch conditions were a pressure of 150 mTorr, a power density of .4 w/cm^2 , a gas flow of 150 sccm each of SF₆ and C₂ClF₆, and an etch time of 20 minutes. There were two sets of samples etched at the same time. One set involved the presence of AZ1470 positive photoresist on the aluminum plate supporting the grid. For the other set, no photoresist was present. Previous work has shown that the erosion of the photoresist in the plasma locally supplies polymer forming

410 Education (Casasas) processors (Casasas)

precursers such as CF₂ to the gas, and results in anisotropic Si etching in regions near photoresist for this etch chemistry.

As soon as the etching was complete, the samples were put into an X-ray photoelectron spectrometer for surface analysis. Time of exposure to the ambient was minimized to the extent possible. The XPS results showed a decrease in the Si surface concentration and a increase the F concentration as one went from open/no resist case to the open/resist case to the grid/no resist case and finally to the grid/resist case. In addition the C 1s peaks showned that the surface concentration of carbon bonded to 2 fluorine atoms followed the same tend as the F concentration for the different samples with a 14x increase in the C-2F concentration between the open/no resist case and the grid/resist case. This indicates an increase in polymer formation as the ion bombardment is decreased and polymer precursers from the resist erosion is introduced. Sputter profile XPS results showed that these surface residues was at most 30 A thick for the open/resist case and less for the other cases.

Etch depth measurements obtained by using a Dektak surface profiler showed that the etch rate decreased by 15x between the open/no resist case and the grid/resist case. When the etch rate is plotted against the C-2F concentration for the different cases, the results show that after a threshold is rearched between the open/no resist and the open/resist cases the etch rate decreases linearly with C-2F concentration. These result indicate that thin (30 A) polymeric layers can block or inhibit etching and are most likely important in many anisotropic etch processes where polymer precursers are present.

Staff: J. D. Shott, J. P. McVittie, K. C. Saraswat, S. H. Goodwin, L. Lewyn, J. D. Plummer, B. Bakoglu.

Related Efforts: Oldham (Berkeley).

References: [Moslehi 87a, Moslehi 87b, Mos 86, Mosl 86, Shah 86, Leeke 86, Uhm 86, McVittie 86a, McVittie 86b]

4.4 Interconnections and Contacts

4.4.1 Objective

With advances in integrated circuit technology, device dimensions are being scaled down and concurrently the chip size and complexity are continuously increasing. requiring closely spaced long interconnection lines with smaller area of the contacts. As a result the RC time delay, the IR voltage drop, the power consumption and cross talk noise associated with the interconnection lines and contacts can become appreciable. Thus, even with very fast devices the overall performance of a large circuit could be seriously affected by the limitations of the interconnections and contacts.

The overall objective of this research is to investigate conducting and insulating materials, fabrication processes, and device structures for multilevel interconnections and contacts in submicron VLSI, so that advances in integrated electronics can continue. Specifically, we are investigating low-pressure CVD of tungsten and tungsten silicide and alloys of aluminum with titanium and other refractory metals to obtain better device structures and to overcome the

problems of VLSI outlined above.

During this period effort has been focussed on Al/Ti/Si metal films, selective low pressure CVD of tungsten and formation of tungsten silicide for gates and interconnections, and technics to measure specific contact resistivity accurately.

Al/Ti/Si Films for Multilevel Interconnections

Layered structures and homogeneous alloy films of Al/Ti/Si were synthesized by sputter deposition and were investigated for use in a VLSI multilevel interconnect technology. Major areas of study include hillock formation, stress measurements during temperature cycling, dry etchability, resistivity before and after annealing, electromigration, film composition and structure, and interlevel shorts. We have demonstrated in this work that aluminum alloyed with silicon and titanium, or layered with titanium offers advantages over current technological materials for interconnections in integrated circuits.

This research has resulted in several specific conclusions pertinent to Al interconnect device technology.

- Alloying of Al/Si with Ti by either layering or as a homogeneous films results in a reduction in hillock density and smooth films at the 2 nm level.
- Alloying of Al with Ti homogeneously does not result in smooth, low resistivity films thus demonstrating the importance of Si.
- Annealing may have a significant effect on the surface morphology through the formation of pillars, which is dependent on the alloying element concentration.
- Both layered and homogeneous Al/Si with Ti structures are dry etchable facilitating processing.
- Resistivities are well within engineering limits for the layered and homogeneous Al/Si-Ti films even after significant annealing, layered films giving the best results.
- As a result of the film uniformity, interlevel shorts are eliminated so that large capacitors and ground planes for multilevel interconnects may be fabricated.
- Electromigration resistance is strongly enhanced in the layered structures in accord with the results of other work.

The impact of this research is that low resistivity, hillock free, dry etchable metal films can be fabricated and used in VLSI multilevel interconnects.

Selective CVD of Tungsten

Selective CVD of tungsten has been investigated to provide a low resistivity shunt over the high resistivity shallow junctions. Two different kinds of shunting layers over the MOS source/drain regions are under consideration. In the first technique W is deposited selectively and the rest of the processing is done in a way to avoid silicidation, i.e., W is not converted to WSi₂. In the other technic after deposition of W annealing is done to obtain silicide. Since both involve selective W, both are selective processes. The potential advantages and disadvantages of each of the two techniques are discussed and then, new process schemes are proposed and will be tested in the near future.

Preliminary work on applying selective W in the source/drain regions of an MOS transistor has been attempted. It is found by doing TEM analysis that if selective W is deposited directly on exposed Si surface, considerable defects are generated at the interface. The influence of these defects on the leakage current of shallow junctions is under investigation.

If the W film is converted to WSi₂ by thermal anneal, then a relatively high temperature cycle is needed to drive down the resistivity to a low value. Also, considerable Si consumption, about 2.5 times the original W thickness occurs during the annealing cycle. Both the high temperature cycle required and the physical consumption of the Si will make ultra-shallow junction more difficult to make in this scheme.

A new scheme is proposed to overcome the W/Si interface problem. A very thin of sputtered W film is deposited over the wafer. Ion implantation is used both to promote a more uniform and controllable silicidation process. Subsequent thermal anneal will convert the W to WSi₂ and activate the implanted dopants at the same time. Unreacted W over SiO₂ is then removed. Next, W is deposited on the WSi₂ film selectively. In this case, only a good ohmic contact between WSi₂ and Si is required, not the low resistivity of the WSi₂ film. Also, during the LPCVD of W, the WSi₂ film will physically separate the reactant gases and the Si substrate, so that the W/Si interface problem may at least be alleviated.

Specific Contact Resistivity Measurement

Specific contact resistivity, ρ_c , is defined to be the ratio of voltage to current density (V/J) for current flowing across a junction of infinitesimal cross-sectional area. It is a physical parameter, typically expressed in Ω -cm² or Ω - μ m², which governs the overall resistance of a contact.

The main objective of this project is to determine the doping density and temperature dependance of ohmic contacts. More specifically, it is desired to quantify the doping and temperature dependance of contact resistivity of various metal-silicon systems and to determine if existing models predict this behavior adequately. Typically, ρ_c increases when temperature is decreased, but for highly doped silicon it is believed that this behavior will reverse, because the current is dominated by tunneling.

Previously, it has been impossible to demonstrate this behavior experimentally, because existing measurement techniques introduce very large error into the estimation of ρ_c . When several devices of differing sizes are used to measure resistivity, the agreement is poor. If the temperature dependance of ohmic contacts is to be determined experimentally, this problem must first be resolved.

In order to explain this anomalous behavior, earlier researchers have hypothesized that ρ_c may be geometry dependant due to macroscopic effects, such as surface pitting along the perimeter of the contact, or other non-uniformities in the interface, but we believe that the error is due to two-dimensional current flow in the test structure, which can not be accounted for by existing one-dimensional models. It is therefore necessary to develop a model which can account for the actual flow of current in the test devices and which can accurately ρ_c from resistance measurements.

To meet this second objective, two-dimensional numerical simulations have been performed to examine the test structures typically used for resistivity measurement. These simulations verify

that the anomalous behavior mentioned above is indeed due to two-dimensional effects and that a two-dimensional model can accurately account for the geometry dependance of such structures. This makes possible the extraction of ρ_c with a much greater degree of confidence. It has been shown that for clean, uniform contacts, ρ_c is indeed a microscopic parameter which does not depend on contact area. Simulations were performed over a very wide variety of geometries and were used to extract the specific resistivity of actual fabricated contacts of Al, CVD W, and PtSi to N⁺ and P⁺ silicon. These contacts showed agreement which was much better than that obtained by previous models.

By examining the fundamental equations governing the current flow through contacts, it was determined that the aforementioned simulations scale by a simple rule. This rule includes scaling of both the contact size and of the sheet resistance of the silicon beneath the contact. By this rule, the simulations made earlier have been normalized, so they may be applied to a wide variety of experimental conditions. A model has been developed which allows extraction of ρ_c via a graphical technique. This model has been used to demonstrate the temperature dependence of ohmic contacts. This work shows that the contact resistivity actually can decrease when temperature is reduced, if the silicon is doped heavily enough. We believe this to be the first time that this effect has been shown.

Staff: Prof. Krishna C. Saraswat, Dr. James P. McVittie, Mr. Donald Gardner, Mr. Han-Chang Wu-Lu, Mr. Man Wong, Mr. William Loh.

Related Efforts: Trotter (Miss. State.).

References: [Loh 86, Schreyer 86, Schrey 86, Gard 87, Saras 86]

energy Datasos of Presentation Francisco

References

[Agarwal 86a] Agarwal, A., Sites, R., Horowitz, M.

ATUM: A New Technique for Capturing Address Traces Using Microcode. In Proc. 13th Sym. on Computer Architecture. IEEE/ACM, Tokyo, Japan, June, 1986.

[Agarwal 86b] Agarwal, A., Horowitz, M., Hennessy, J.

An Analytical Cache Model.

Technical Report CSL-TR-86-304, Computer Systems Laboratory, Stanford University, September, 1986.

[Agarwal 87] Agarwal, A., Chow, P., Horowitz, M., Acken, J., Salz, A., and hennessy, J.

On-Chip Instruction Caches for High Performance Processors.

In Advanced Research in VLSI. Stanford University, Stanford, Ca., March, 1987.

[Buehler 80] Martin G. Buehler.

The Use of Electrical Test Structure Arrays for Integrated Circuit Process Evaluation.

J. Electrochemical Society 127:2284-2290, Oct., 1980.

[Chae 86] S.I. Chae, J.T. Walker, D.H. Dameron, C.C. Fu, and J.D. Meindl.

Template-set approach to VLSI pattern inspection.

Automated Inspection and Measurement, SPIE J. 730, 1986.

[Chen 85] Ihao Chen.

A Methodology for Optimal Test Structure Design.

Master's thesis, Carnegie Mellon University, May, 1985.

[Chow 83] Chow, F.

A Portable Machine-Independent Gloabl Optimizer - Design and

Measurements.

PhD thesis, Stanford University, December, 1983.

[Chow 85] Chow, P., Horowitz, M.

The MIPS-X Microprocessor.

In Proc. WESCON 85. IEEE, San Francisco, 1985.

[Chow 86] Chow, P.

MIPS-X Instruction Set and Programmer's Manual.

Technical Report 86-289, Computer Systems Laboratory, Stanford University, May, 1986.

[ChowHenn 84] Chow, F.C. and Hennessy, J.L.

Register allocation by priority-based coloring.

In Proc. of 1984 Compiler Construction Conference, pages. ACM, Montreal, Canada, June, 1984.

[Chu 86] Chu, C.Y., Horowitz, M.

Charge Sharing Models for MOS Circuits.

In Intl. Conference on Computer-Aided Design. IEEE, November, 1986.

[Gard 87] D. Gardner, J.D. Meindl and K.C. Saraswat.

Interconnection and Electromigration Scaling Theory.

March 1987.

IEEE Trans. Electron Devices.

[Gross 83] Gross, T.R.

Code Optimization of Pipeline Constraints.
PhD thesis, Stanford University, August, 1983.

[Hall 67] P. M. Hall.

Resistance Calculations for Thin Film Patterns.

Thin Solid Films 1:277-295, 1967.

[Hennessy 84] Hennessy, J.

VLSI Processor Architecture.

IEEE Trans. on Computers C-33(12):1221-1246, December, 1984.

25th anniversary issue; invited paper.

[Hennessy 86] Hennessy, J. and Horowitz, M.

An Overview of the MIPS-XMP Project.

Technical Report 86-300, Computer Systems Laboratory, Stanford University,

April, 1986.

[Horowitz 87] Horowitz, M, et al.

A 12 MIPS Microprocessor with On-Chip Cache.

In International Solid State Circuits Conference. IEEE, February, 1987.

[Leeke 86] S.D. Leeke, D.K.Y. Liu, and J. P. McVittie.

Plasma Mode Trench Etching with Direct Hydrocarbon Injection.

In 1986 Symposium Proceedings, pages 21. Materials Research Society,

1986.

[Liu 86] Liu, D. and McCluskey, E.

Design of CMOS VLSI Circuits for Testability.

In Proc. 1986 Custom Integrated Circuits Conference. Rochester, NY, 1986.

[Liu 87] Liu, D. and McCluskey, E.

A VLSI CMOS Circuit Design Technique to Aid Test Generation.

In Proc. Phoenix Conf. on Computers and Communications. Scottsdale, Az.,

1987. Ta annsa

To appear.

[Loh 86] W.M. Loh, P.J. Wright, T.A. Schreyer, S.E. Swirhun, K.C. Saraswat and J.C.

Meindl.

The Sidewall Resistor - A Novel Test Structure to Reliably Extract Specific

Contact Resistivity.

IEEE Trans. Electron Devices EDL-7(8):447-449, August, 1986.

[lukaszek 86] W. Lukaszek, W. Yarbrough, J. T. Walker, and J. Meindl.

CMOS Test Chip Design for Process Problem Debugging and Yield

Prediction Experiments.

Solid-State Technology, March, 1986.

A.M. McCarthy, W. Lukaszek, and J.D. Meindl. [mccarthy 86]

A New Ion Implant Monitor Electrical Test Structure.

In VLSI Workshop on Test Structures, pages 78-80. IEEE, Long Beach, CA, February, 1986.

McFarling, S., Hennessy, J. [McFarling 86]

Reducing the Cost of Branches.

In Proc. 13th Sym. on Computer Architecture. IEEE/ACM, Tokyo, Japan, June. 1986.

J.P. McVittie and T.A. Lin. [McVittie 86a]

Ion Suppression for Studying Etch Inhibitor Layers.

In Extended Abstracts, pages 490. Electrochemical Society, October, 1986.

[McVittie 86b] J.P. McVittie, T.A. Lin, and A.J. Bariya.

Ion Suppression for Studying A New Method for Analyzing Thin Wall

Inhibitor Layers.

In Symposium Proceedings. Materials Research Society, December, 1986. To Be Published.

[Miyamoto 87] J. Miyamoto, M. Horowitz.

A Single Chip LSI High Speed Functional Tester.

In International Solid State Circuits Conference. IEEE, February, 1987.

M.M. Moslehi, S.C. Shatas and K.C. Saraswat. [Mos 86]

Rapid Thermal Oxidation of Silicon.

Submitted to the Fifth International Symposium on Silicon Materials Science

and Technology, May, 1986.

Boston.

[Mosl 86] M.M. Moslehi.

Thermal and Plasma Nitridation of Si and SiO₂ for Ultrathin gate Insulators of

Presented at the Automation '86 International High Technol. Convention and Exhibition, March, 1986.

Invited Paper, Houston.

[Moslehi 87a] Mehrdad M. Moslehi, Steven C. Shatas, and Krishna C. Saraswat.

> Interfacial and Breakdown Characteristics of MOS Devices with Rapidly Grown Ultrathin SiO₂ Gate Insulators.

Applied Physics Letters, 1987.

To be published.

[Moslehi 87b] Mehrdad M. Moslehi.

Radip Thermal/Plasma Processing for in-situ Dielectric Engineering.

In 1987 Symposium on Rapid Thermal Processing of Electronics Materials.

Marterials Research Society, Anaheim, CA, April, 1987.

To be published.

[Saras 86] K. Saraswat, W. Loh, T. Schreyer and S. Swirhun.

Measurement and Extraction of Specific Contact Resistivity.

In Proceedings 3rd International IEEE VLSI Multilevel Interconnection

Conference, pages 385-395. IEEE, Santa Clara, June, 1986.

Sarkar, V., Hennessy, J. [Sarkar 86a]

Compile-time Partitioning and Scheduling of Parallel Programs.

In Sym. on Compiler Construction. ACM, Palo Alto, Ca., June, 1986.

[Sarkar 86b] Sarkar, V. and Hennessy, J.

Partitioning Parallel Programs for Macro-Dataflow.

In Sym on LISP and Functional Programming. ACM, Boston, Mass., August,

1986.

T. A. Schreyer and K.C. Saraswat. [Schrey 86]

A Two-Dimensional Analytical Model for the Cross-Bridge Kelvin Resistor.

IEEE Electron Devices Letters EDL-7(11), December, 1986.

T. Schreyer, S. Swirhun, W. Loh, K. Saraswat and R. Swanson. [Schreyer 86]

Comparison of Test Structures used for the Measurement of Low Resistive

Metal-Semiconductor Contacts.

IEEE VLSI Workshop on Test Structures: 7-23, February, 1986.

Long Beach, CA.

N. Shah, J. McVittie, N. Sharif, J. Nulman, and A. Gat. [Shah 86]

Characterization of PSG Films Reflowed in Steam Using Rapid Thermal

Processing.

In 1986 Symposium Proceedings, pages 233. Materials Research Society,

1986.

SteenKiste, P. and Hennessy, J. [Steenkiste 86]

LISP on a Reduced-Instruction-Set Processor.

In Sym. on LISP and Functional Programming. ACM, Boston, Mass.,

August, 1986.

N.L. Turner, M. Current, T.C. Smith, and D. Crane. [Turner 85]

Effects of Planar Channeling using Modern Ion-Implantation Equipment.

Solid State Technology: 163-172, February, 1985.

[Uhm 86] K.S. Uhm, M.R. Kump, J. P. McVittie, and R.W. Dutton.

Kinetic Modeling and Measurement of Active Species Distribution During

Dry Etching.

In 1986 Symposium Proceedings, pages 237. Materials Research Society,

1986.

[Yamaguchi 86] R. Yamaguchi, Komatsu, K., Moriya, S., Harada, K.

Integrated Electrical Vernier to Measure Registration Accuracy.

IEEE Electron Device Letters EDL-7 No. 8:463, 464, August, 1986.

[yarbrough 86] W. J. Yarbrough, W. Lukaszek, and J.D. Meindl.

VLSI Process Problem Diagnosis and Yield Prediction: A Comprehensive

Test Structure and Test Chip Design Methodology.

In VLSI Workshop on Test Structures, pages 294-304. IEEE, Long Beach,

CA, February, 1986.

Publications

Agarwal, A., Sites, R., Horowitz, M. [Agarwal 86a]

ATUM: A New Technique for Capturing Address Traces Using Microcode. In Proc. 13th Sym. on Computer Architecture. IEEE/ACM, Tokyo, Japan, June, 1986.

[Agarwal 86b] Agarwal, A., Horowitz, M., Hennessy, J.

An Analytical Cache Model.

Technical Report CSL-TR-86-304, Computer Systems Laboratory, Stanford University, September, 1986.

Agarwal, A., Chow, P., Horowitz, M., Acken, J., Salz, A., and hennessy, J. [Agarwal 87]

On-Chip Instruction Caches for High Performance Processors.

In Advanced Research in VLSI. Stanford University, Stanford, Ca., March, 1987.

S.I. Chae, J.T. Walker, D.H. Dameron, C.C. Fu, and J.D. Meindl. [Chae 86]

Template-set approach to VLSI pattern inspection.

Automated Inspection and Measurement, SPIE J. 730, 1986.

[Chow 85] Chow, P., Horowitz, M.

The MIPS-X Microprocessor.

In Proc. WESCON 85. IEEE, San Francisco, 1985.

[Chow 86] Chow, P.

MIPS-X Instruction Set and Programmer's Manual.

Technical Report 86-289, Computer Systems Laboratory, Stanford University, May, 1986.

Chu, C.Y., Horowitz, M. [Chu 86]

Charge Sharing Models for MOS Circuits.

In Intl. Conference on Computer-Aided Design. 1EEE, November, 1986.

D. Gardner, J.D. Meindl and K.C. Saraswat. [Gard 87]

Interconnection and Electromigration Scaling Theory.

March 1987.

IEEE Trans. Electron Devices.

Horowitz, M, et al. [Horowitz 87]

A 12 MIPS Microprocessor with On-Chip Cache.

In International Solid State Circuits Conference. IEEE, February, 1987.

S.D. Leeke, D.K.Y. Liu, and J. P. McVittie. [Leeke 86]

Plasma Mode Trench Etching with Direct Hydrocarbon Injection.

In 1986 Symposium Proceedings, pages 21. Materials Research Society,

1986.

Liu, D. and McCluskey, E. [Liu 86]

Design of CMOS VLSI Circuits for Testability.

In Proc. 1986 Custom Integrated Circuits Conference. Rochester, NY, 1986.

[Liu 87] Liu, D. and McCluskey, E.

A VLSI CMOS Circuit Design Technique to Aid Test Generation.

In Proc. Phoenix Conf. on Computers and Communications. Scottsdale, Az., 1987.

То арреаг.

[Loh 86] W.M. Loh, P.J. Wright, T.A. Schreyer, S.E. Swirhun. K.C. Saraswat and J.C.

Meindl.

The Sidewall Resistor - A Novel Test Structure to Reliably Extract Specific

Contact Resistivity.

IEEE Trans. Electron Devices EDL-7(8):447-449, August, 1986.

[McFarling 86] McFarling, S., Hennessy, J.

Reducing the Cost of Branches.

In Proc. 13th Sym. on Computer Architecture. IEEE/ACM, Tokyo, Japan,

June, 1986.

[McVittie 86a] J.P. McVittie and T.A. Lin.

lon Suppression for Studying Etch Inhibitor Layers.

In Extended Abstracts, pages 490. Electrochemical Society, October, 1986.

[McVittie 86b] J.P. McVittie, T.A. Lin, and A.J. Bariya.

Ion Suppression for Studying A New Method for Analyzing Thin Wall

Inhibitor Layers.

In Symposium Proceedings. Materials Research Society, December, 1986.

To Be Published.

[Mos 86] M.M. Moslehi, S.C. Shatas and K.C. Saraswat.

Rapid Thermal Oxidation of Silicon.

Submitted to the Fifth International Symposium on Silicon Materials Science

and Technology, May, 1986.

Boston.

[Mosl 86] M.M. Moslehi.

Thermal and Plasma Nitridation of Si and SiO₂ for Ultrathin gate Insulators of

MOS VLSI.

Presented at the Automation '86 International High Technol. Convention and

Exhibition, March, 1986.

Invited Paper, Houston.

[Moslehi 87a] Mehrdad M. Moslehi, Steven C. Shatas, and Krishna C. Saraswat.

Interfacial and Breakdown Characteristics of MOS Devices with Rapidly

Grown Ultrathin SiO₂ Gate Insulators.

Applied Physics Letters, 1987.

To be published.

[Moslehi 87b] Mehrdad M. Moslehi.

Radip Thermal/Plasma Processing for in-situ Dielectric Engineering.

In 1987 Symposium on Rapid Thermal Processing of Electronics Materials.

Marterials Research Society, Anaheim, CA, April, 1987.

To be published.

[Saras 86] K. Saraswat, W. Loh, T. Schreyer and S. Swirhun.

Measurement and Extraction of Specific Contact Resistivity.

In Proceedings 3rd International IEEE VLSI Multilevel Interconnection Conference, pages 385-395. IEEE, Santa Clara, June, 1986.

[Sarkar 86a] Sarkar, V., Hennessy, J.

Compile-time Partitioning and Scheduling of Parallel Programs.

In Sym. on Compiler Construction. ACM, Palo Alto, Ca., June; 1986.

[Sarkar 86b] Sarkar, V. and Hennessy, J.

Partitioning Parallel Programs for Macro-Dataflow.

In Sym on LISP and Functional Programming. ACM, Boston, Mass., August, 1986.

[Schafft 86] H. Schafft and J.D. Shott.

An Interlaboratory Comparison of Electromigration Test Methods.

IEEE Transactions on Electron Devices, December, 1986.

To be published.

[Schrey 86] T. A. Schreyer and K.C. Saraswat.

A Two-Dimensional Analytical Model for the Cross-Bridge Kelvin Resistor.

IEEE Electron Devices Letters EDL-7(11), December, 1986.

[Schreyer 86] T. Schreyer, S. Swirhun, W. Loh, K. Saraswat and R. Swanson.

Comparison of Test Structures used for the Measurement of Low Resistive

Metal-Semiconductor Contacts.

IEEE VLSI Workshop on Test Structures: 7-23, February, 1986.

Long Beach, CA.

[Shah 86] N. Shah, J. McVittie, N. Sharif, J. Nulman, and A. Gat.

Characterization of PSG Films Reflowed in Steam Using Rapid Thermal

Processing.

In 1986 Symposium Proceedings, pages 233. Materials Research Society,

1986.

[Steenkiste 86] SteenKiste, P. and Hennessy, J.

LISP on a Reduced-Instruction-Set Processor.

In Sym. on LISP and Functional Programming. ACM, Boston, Mass.,

August, 1986.

[Uhm 86] K.S. Uhm, M.R. Kump, J. P. McVittie, and R.W. Dutton.

Kinetic Modeling and Measurement of Active Species Distribution During

Dry Etching.

In 1986 Symposium Proceedings, pages 237. Materials Research Society,

1986.